Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Cheryl Ann Szkudlarek

Entitled
Multivariate Statistical Methods Applied to the Analysis of Trace Evidence

For the degree of Master of Science

Is approved by the final examining committee:

John V. Goodpaster
Chair

Jay A. Siegel

Robert Minto

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): John V. Goodpaster

Approved by: Christine Picard 03/06/2012
Head of the Graduate Program Date

Graduate School Form 20
(Revised 9/10)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:

Multivariate Statistical Methods Applied to the Analysis of Trace Evidence

For the degree of Master of Science

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Cheryl Ann Szkudlarek

Printed Name and Signature of Candidate

03/06/2012

Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

MULTIVARIATE STATISTICAL METHODS APPLIED TO THE ANALYSIS OF TRACE EVIDENCE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Cheryl Ann Szkudlarek

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2012

Purdue University

Indianapolis, Indiana

For my family, you are always there to aid and support me especially when stress levels were high. Mom and dad, you believe in and encourage me even though my research may be beyond your knowledge. I truly and deeply appreciate everything you have done for me. My sisters, you help me through tough situations and are always there for me. Lastly, for those who have believed in me and helped me throughout the years.

ACKNOWLEDGMENTS

I would like to give a special thanks to Dr. John V. Goodpaster, my research advisor and mentor, for accepting me and giving me the opportunity to fulfill my academic goals. I would also like to thank all faculty members who have aided me throughout my studies. The idea for the red cotton fiber study was suggested by the Microtrace Unit at the Indiana State Police Laboratory, so I sincerely thank you. Thank you Testfabrics, Inc. and Dr. Stephan Morgan at the University of South Carolina for providing the dyed cotton exemplars. I want to thank Elisa Liszewski Pozywio for collecting all of the fiber absorbance data. I would also like to thank the Federal Bureau of Investigation Laboratory and the Alcohol, Tobacco, Firearms, and Explosives Laboratory for allowing me to assist them in their most recent electrical tape studies. Huge thanks to the Midwestern Forensic Resource Center (MFRC) and the National Institute of Justice (NIJ) for the financial support to make the projects possible. Lastly, I want to thank all friends, family, and teachers who have been a positive influence in my studies.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF ABBREVIATIONS

^{o}	degree
x	magnification
AHC	Agglomerative Hierarchical Clustering
ANOVA	Analysis of Variance
BBP	Butyl benzyl phthalate
CE	Capillary Electrophoresis
CI	Colour Index
CV	canonical variate
DA	Discriminant Analysis
DART	Direct Analysis in Real Time
DEP	Diethyl phthalate
DMP	Dimethyl phthalate
DNP	Dinonyl phthalate
DOA	Dioctyl adipate
DOP	Dioctyl phthalate
FTIR	Fourier Transform Infrared Spectroscopy
GC	Gas Chromatography
HPLC	High Performance Liquid Chromatography
ISP	Indiana State Police
LC	Liquid Chromatography
m/z	Mass per charge ratio
MS	Mass Spectrometry
MSP	Microspectrophotometry
MSS	Mean Sum of Squared

nm	nanometer
PC	principal component
PCA	Principal Components Analysis
PLM	Polarized Light Microscopy
PVC	polyvinyl chloride
Py	Pyrolysis
SEM-EDS	Ccanning Electron Microscopy-Energy Dispersive Spectrometry
SWGMAT	Scientific Working Group on Materials Analysis
TLC	Thin Layer Chromatography
TOTM	Trioctyl trimellitate
UV	ultraviolet
Vis	visible
XRD	X-Ray Diffraction
XRF	X-Ray Fluorescence

ABSTRACT

Szkudlarek, Cheryl Ann. M.S., Purdue University, May 2012. Multivariate Statistical Methods Applied to the Analysis of Trace Evidence. Major Professor: John V. Goodpaster.

The aim of this study was to use multivariate statistical techniques to: (1) determine the reproducibility of fiber evidence analyzed by MSP, (2) determine whether XRF is an appropriate technique for forensic tape analysis, and (3) determine if DART/MS is an appropriate technique for forensic tape analysis. This was achieved by employing several multivariate statistical techniques including agglomerative hierarchical clustering, principal component analysis, discriminant analysis, and analysis of variance. First, twelve dyed textile fibers were analyzed by UV-Visible MSP. This analysis included an inter-laboratory study, external validations, differing preprocessing techniques, and color coordinates. The inter-laboratory study showed no statistically significant difference between the different instruments. The external validations had overall acceptable results. Using first derivatives as a preprocessing technique and color coordinates to define color did not result in any additional information. Next, the tape backings of thirty-three brands were analyzed by XRF. After chemometric analysis it was concluded that the 3M tapes with black adhesive can be classified by brand except for Super 33+ (Cold Weather) and Super 88. The colorless adhesive tapes were separated into two large groups which were correlated with the presence of aluminosilicate filler. Overall, no additional discrimination was seen by using XRF compared to the traditional instrumentation for tape analysis previously published. Lastly, the backings of

eighty-nine brands of tape were analyzed by DART/MS. The analysis of the black adhesive tapes showed that again discrimination between brands is possible except for Super 33+ and Super 88. However, now Tartan and Temflex have become indistinguishable. The colorless adhesive tapes again were more or less indistinguishable from one another with the exception of Tuff Hand Tool, Qualpack, and a roll of 3M Tartan, which were found to be unique. It cannot be determined if additional discrimination was achieved with DART/MS because the multivariate statistical techniques have not been applied to the other instrumental techniques used during tape analysis.

CHAPTER 1. INTRODUCTION

1.1. Chemometrics Applied to Forensic Science

Chemometrics is the application of multivariate statistics to chemical data (e.g. absorbance spectra or chromatograms). Multivariate statistics is the application of statistics to a data set containing more than one variable. The use of chemometrics in forensic science is growing in popularity. Forensic scientists often find themselves comparing spectra or chromatograms, interpreting differences, and identifying patterns. Typically, such analysis relies on visual comparisons and inspection performed by the analyst. The problems start arising when the data set contains more than two or three variables (dimensions). A human's ability to perceive beyond two- or three-dimensions is weak. Therefore, chemometrics make the task of examining large data sets for patterns more accurate, objective, and manageable. Since its inception into the forensic discipline it has been applied to accelerants, document examination, fibers, glass, gunpowder, inks, and paint.[1]

As mentioned earlier, current techniques rely upon the analyst's perception when determining whether an unknown sample and known samples have a common origin. Therefore, there is limited statistical basis to determine the evidentiary value of the analysis. This is a growing issue for the forensic laboratory as addressed in the National Academy of Sciences (NAS) report on strengthening forensic science.[2] The issues that can be addressed with chemometrics are Recommendations 3 and 5. Recommendation 3 addresses the accuracy, reliability, and validity of trace evidence analysis. Recommendation 5 addresses human observer bias and sources of human error that occur during trace evidence analysis.[2] Overall, these issues affect the admissibility of trace evidence analysis into the court room. Therefore,

chemometrics can help resolve the conditions for admissibility as stated in the case of *Daubert v. Merrell Dow Pharmaceuticals*.[3]

It can be generalized that chemometrics methods are utilized to reduce the dimensionality, sort, and group the variables of large data sets. Also, chemometrics is used to investigate the dependence or correlation of variables, predicting placement of unknown samples, or constructing hypotheses.[4] Overall, chemometrics techniques are a great tool when large data set have been acquired.[1] Four chemometrics techniques were utilized in this study: Agglomerative Hierarchical Clustering (AHC), Principal Component Analysis (PCA), Discriminant Analysis (DA), and Analysis of Variance (ANOVA) after preprocessing. The theory behind PCA has been around for over a century, when it was first introduced by Pearson in 1901. However, the algorithm was not introduced until 1933 by Hotelling. In 1936, Fisher developed DA and Mahalanobis developed the distance measurement utilized in DA. Lastly, ANOVA was developed by Ronald Fisher in 1918. However, it has not been considered in forensic science until recently.[1]

1.2. Preprocessing Techniques

Preprocessing is defined as the mathematical alteration of the data set before analysis. These techniques help eliminate or reduce random or systematic sources of variation that may be contained within the data set, which may hinder the overall interpretation during analysis. Care must be taken when choosing the appropriate technique to be applied because preprocessing changes the data set. When an inappropriate technique has been employed important variations can be masked or removed from the data set.[5]

There are two ways to preprocess a data set; applying the process to either the samples or the variables. Three common techniques are available to preprocess the samples: smoothing, baseline correction, and normalization. Smoothing can reduce the amount of unnecessary noise and increase the signal-to-noise ratio; while assuming that the noise occurs at a higher frequency than

the signal of interest. This is achieved one of five ways: mean smoother, running mean smoother, running median smoother, running polynomial smoother, or Fourier filter smoother. The most commonly used is the running polynomial smoother. Mean smoother removes data points from the sample vector while running mean and running median smoothers cause "end effects" because the ends cannot be smoothed in the same manner as the other points.[5] Running polynomial smoother fits a low-order polynomial to the points in the window using the Gorry-modified Savitzky-Golay algorithm.[1] These methods all require a window width, where the points inside the window are considered during the calculations. Therefore, the window width chosen is very important. The last method the Fourier filter smoother, transforms the signal, weights each transformed point, and transforms them back into the units of the original data. This method does not involve a window width but is dependent on the apodization function chosen.[5]

Baseline correction accounts for systematic variation and varying background levels that may cause confusion during interpretation.[1,5] The sample vector can be written as a function equal to the signal of interest plus some background features, as presented as Equation 1.1, where $\tilde{r}(x)$ is the signal of interest and the remaining coefficients are the baseline features.[5]

$$r(x) = \tilde{r}(x) + \alpha + \beta x + \gamma x^2 + \delta x^3 + \cdots \qquad \text{Equation 1.1}$$

Therefore, the baseline can be accounted for by estimating the necessary coefficients and subtracting them from each element in the sample. If an offset baseline is present then Equation 1.1 becomes simplified as the signal of interest and the first coefficient term, α. Then the baseline can be accounted for by estimating α and subtracting it. When a linearly sloping baseline is present then Equation 1.1 becomes the signal of interest plus the first two coefficient terms, α and β. Therefore, a line is estimated defined by α and β which is subtracted from the sample vector. Baseline correction can also be accomplished by replacing each point with its derivative. The first derivative of Equation 1.1 is presented as Equation 1.2, where $\tilde{r}'$(x) represents the first derivative of the signal of interest.[5]

$$r'(x) = \tilde{r}'(x) + 0 + \beta + 2\gamma x + 3\delta x^2 + \cdots \quad \text{Equation 1.2}$$

For the offset baseline, the first derivative completely removes the baseline features because α becomes zero. With successive derivatives all the higher-order terms will be removed. The second derivative is presented as Equation 1.3, where $\tilde{r}''$(x) is the second derivative of the signal of interest.[5]

$$r''(x) = \tilde{r}''(x) + 0 + 0 + 2\gamma + 6\delta x + \cdots \quad \text{Equation 1.3}$$

There are many methods to determine the derivatives: running simple difference, running mean difference, and Gorry-Savitzky-Golay algorithm. The running simple difference estimates the derivative by calculating the difference between adjacent points present in the window. Running mean difference calculates the difference between the individual points present in the selected window. The Gorry-Savitzky-Golay algorithm fits a polynomial to a region of the vector and replaces the point with its derivative, without the "end effects".[5]

Systematic variations arise due to sample amount, concentration, sample size, and instrument responses. These variations can be eliminated by normalization which places all the samples on the same scale. Normalizing the data set to unit area is achieved by dividing each variable in the sample by the sum of the absolute value of all variables in the sample. A second approach is to normalize to unit length; each variable in the sample is divided by the square root of the sum of all values squared. The last approach involves normalizing to the maximum value; each variable in the sample is divided by the maximum value among all variables. Equation 1.4 presents the equation for normalizing to unit length, where x_i is the ith variable of sample x.[5,6]

$$x_{norm} = \frac{x_i}{\sqrt{\sum_{i=1}^{n} x_i^2}} \quad \text{Equation 1.4}$$

Mean centering and variable weighting are among the techniques utilized to preprocess the variables of a data set. Mean centering accounts for undue influences from the points farthest from the original origin by shifting the origin of the coordinate system to the center of the data.[1,7] The mean of each variable is subtracted from the elements of the variable.[5] Unfortunately, mean centering

loses information about the origin of the factor space, relative eigenvalue magnitudes, and relative error.[7] Therefore, mean centering may not always be appropriate. If concerned about the variability above a baseline and not around a mean; as in signal analysis, then mean centering is not an appropriate preprocessing technique.[8] Otherwise, mean centering seldom hinders and often helps.[5,6]

Variable weighting places emphasis on some variables over others increasing their impact during analysis. There are four methods developed for variable weighting: a priori information weighting, variable selection, variance scaling, and autoscaling. A priori information weighting and variable selection methods both require accurate information/knowledge of the uncertainties, analytical instruments, or chemistry of the system. Variance scaling eliminates the units associated with the variables by equalizing the variances of the variables. Each element of the variable is divided by the standard deviation of that variable. This is beneficial when different units are within the data set. Lastly, autoscaling is a combination of mean centering and variance scaling.[5,6] Autoscaling essentially places the variables on a more equal footing by increasing or decreasing the influence on the calibration of each variable. However, autoscaling can have a negative impact on the accuracy and precision of a calibration if applied when not necessary.[7] This method is recommended when large differences in variance are present within the variables or the data set contains different units.[1,6] Overall, preprocessing techniques should be applied with caution.

1.3. Agglomerative Hierarchical Clustering

Cluster analysis is an unsupervised pattern recognition method that allows an analyst to determine the relationship between samples of a data set.[9] Other uses include discovering the intrinsic structure or underlying behavior of the data set, checking the validity of the data set, detecting outliers, and reducing or evaluating the dimensionality of the data set.[4,10] There a two typical types of

cluster analysis: partitioning and hierarchical techniques. With partitioning the samples are moved between the clusters until stability is reached with the predetermined number of clusters.[11] The remainder of this discussion will focus on the hierarchical techniques. Two common techniques are agglomerative hierarchical clustering (AHC) and divisive hierarchical clustering (DHC). These techniques begin by calculating the distance of each sample to all the remaining samples. In DHC all samples begin in a single group and are divided into smaller clusters (with the sample with the largest distance separating from the group first) until all samples are their own group.[4,11] AHC is the complete opposite to DHC in that all samples begin as their own group and the samples with the closest distance are clustered together until a single cluster remains.[4,5,11]

The distance between samples can be calculated either as similarity or dissimilarity. The first and most common method to determine the dissimilarity between two samples is Euclidean distance or ruler distance. The calculation for Euclidean distance is based on the Pythagorean Theorem and represented in Equation 1.5 where $d_{x,y}$ is the distance between the points and (x-y)' is the transpose of the matrix (x-y).[1,4,8]

$$d_{x,y} = \sqrt{(x-y)'(x-y)} \qquad \text{Equation 1.5}$$

The smaller the $d_{x,y}$ value the more similar the samples. A second mathematical calculation for dissimilarity that can be utilized is Manhattan distance or city-block distance. Manhattan distance utilizes the sum of the sides of the triangle to calculate distance whereas Euclidean distance utilized the hypotenuse; therefore the distance measured by Manhattan distance will always be slightly larger than the distance measured by Euclidean distance.[8,9] The Manhattan distance is presented in Equation 1.6.[12]

$$d_{x,y} = \sum_i |x_i - y_i| \qquad \text{Equation 1.6}$$

The last method to be discussed to calculate dissimilarity is Mahalanobis distance which unlike the other two methods takes into account that some variables in the data set may be correlated. It is calculated according to Equation 1.7 where C is the variance-covariance matrix.[8,9]

$$d_{x,y} = \sqrt{(x-y)C^{-1}(x-y)'} \qquad \text{Equation 1.7}$$

The Mahalanobis distance is similar to Euclidean distance with the exception that Mahalanobis distance utilizes the inverse of the variance-covariance matrix as a scaling factor. A disadvantage is that this method is inappropriate if the number of variables exceeds the number of samples. A common method to measure the similarity between samples is by calculating the sample correlation coefficient. The larger the value for the coefficient the more similar the samples are to one another. There are many other methods to measure the similarity that will not be discussed here.[8,9]

Once the distance has been calculated, the samples are linked to one another. There are many methods that can be employed to link the samples. The simplest is the nearest neighbor or single linkage which links clusters based on the distance between the two closest samples of each respective group. A second approach is the opposite of the nearest neighbor which is farthest neighbor or complete linkage. In this method the linkage is based on the distance between the two farthest members of each group.[10] A third method and the one used in this study is Ward's method. This method seeks to minimize the error sum of squares which is determined by measuring the sum of squared deviation of each samples from the mean of the cluster. Therefore, this method must consider all possible pair of clusters during every step and the two clusters whose union would impact the sum of squares the least are linked.[4,12] Other methods that can be employed are weighted or un-weighted pair group average, centroid linkage, and median linkage.[12]

These linkage methods result in a two-dimensional plot called a dendrogram; an example shown in Figure 1.3.1. The dendrogram portrays the steps taken during the clustering process.[4,5] The dendrogram is a form of a tree diagram where the samples are the leaves and the distance values are the branches. A node indicates the linkage of two branches.[12] Referring to Figure 1.3.1, data provided by Dr. John Goodpaster, the shorter the branch the more similar the cities to one another; e.g. the cities within the same state were most

similar to one another indicated by the short distance. Once the cities were clustered the states were clustered together; e.g. Maryland and Massachusetts. The vertical dotted line is the truncation line which indicates how many grouping the data set contains; groupings to the right are significant. This line is determined based on the distances and the first major increase in the branches; therefore, there are three significant grouping within the data set. A class centroids plot, which portrays the centroid for each group, and a central objects plot, which portrays the sample that is the closest to the centroid of the designated group are provided to show what the characteristics of the groups are.[12]

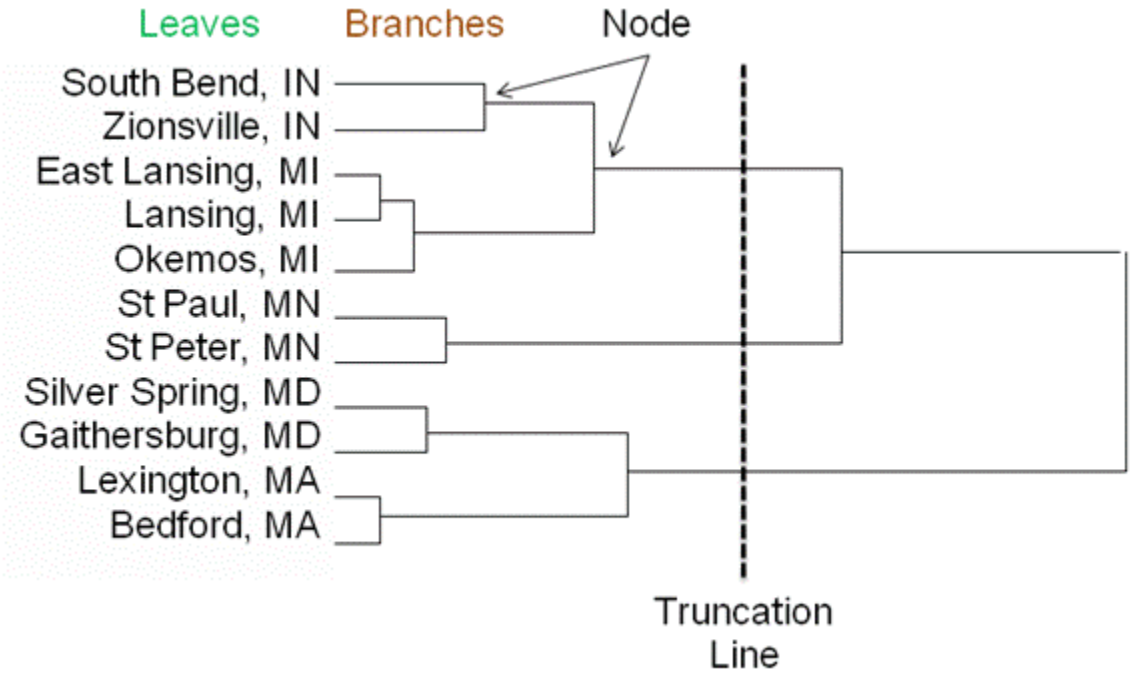


Figure 1.3.1 Example of a AHC dendrogram (courtesy of John Goodpaster)

In summary, AHC is an appropriate technique to determine the relationship between samples in a large data set whether that is similarity or dissimilarity. The dendrogram is a simple representation of this relationship but does not indicate which variable(s) contribute the most to the relationship. However, AHC is sensitive to outliers and noise. Therefore, the presence of an outlier would be apparent in the dendrogram. Cluster analysis, including the

AHC method, has previously been applied to the analysis of cotton fibers (white)[13], counterfeit coins[14], electrical tape[15,16], hair dye[17], illicit drugs[18], lighter fuel[19], lubricating oil[20], paint[21], pen ink[22], polymers[23], and soil[24-25].

1.4. Principal Component Analysis

Principal Component Analysis (PCA) is first and foremost a data dimensionality reduction technique. PCA is used to visualize and identify if variable to variable correlation exists and to portray the data in two- or three-dimensions to allow the identification of groupings.[12] This is achieved by identifying a new set of axes called principal components (PCs) or factors that are linear combinations of the original variables and explain all or almost all of the total variance.[26] Therefore, the first PC (represented in Equation 1.8 as a linear function) is calculated so that it explains the maximum amount of variation of the data set given the constraint in Equation 1.9 where X_n represent the original variables and a_{mn} are constants.

$$PC1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \qquad \text{Equation 1.8}$$

$$a_{11}^2 + a_{12}^2 + \cdots + a_{1n}^2 = 1 \qquad \text{Equation 1.9}$$

The second principal component (represented as a linear function in Equation 1.10) is chosen so that it captures as much as possible of the remaining total variance given the constraint of Equation 1.11.[1,11]

$$PC2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \qquad \text{Equation 1.10}$$

$$a_{21}^2 + a_{22}^2 + \cdots + a_{2n}^2 = 1 \qquad \text{Equation 1.11}$$

Successive PCs are calculated in the same way and will explain a portion of the remaining variance. Each PC must be orthogonal to the previous one. The possible number of PCs is either the number of the samples or the variables whichever is smaller.[1,5] Since, the majority of the variance is explained by the first few PCs, those that explain relatively small amounts of the total variance can be considered negligible. Therefore, the original data set is now presented with fewer dimensions but retains most of the measurement information. This can be

visually represented as an observations plot, as shown in Figure 1.4.1, which indicates how samples are related by plotting the factor scores of one PC against the factor scores of another. Factor scores are the new coordinates that were determined based on the new set of axes. Samples that lie close to one another in this plot are more closely related than samples that lie on opposite sides of the axes.[5] Referring to Figure 1.4.1 the cities within the states Massachusetts, Maryland, Michigan, and Minnesota are more similar to one another than cities from the other states. The cities from Indiana do not appear to be similar indicated by the gap.

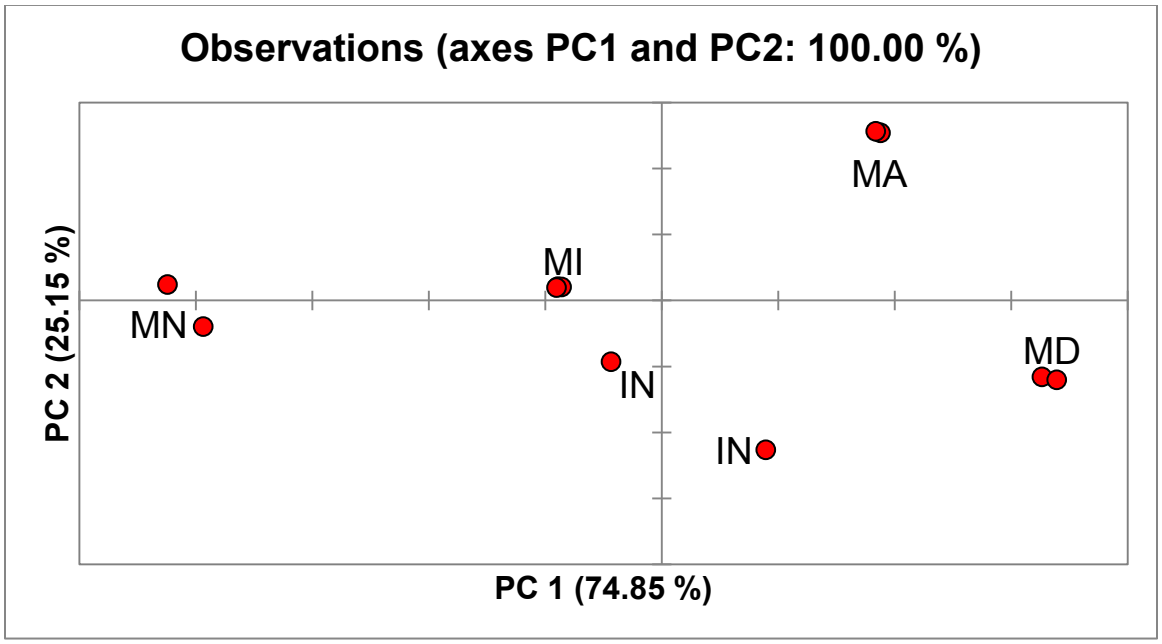


Figure 1.4.1 Example of a PCA observations plot (courtesy of John Goodpaster)

The contribution of each variable to the new principal component defined as the cosine of the angle between the variable axis and the principal component axis. These cosine values are called factor loadings, which can have the value of ±1. When the value is positive the variable and PC are positively correlated. When the value is close to zero then no correlation exists between the variable

and the PC. Lastly, if the value is negative then a negative correlation exists between the variable and the PC.[5]

Rotating the principal components can make difficult interpretation of factor loadings easier. The two types of rotational methods are orthogonal and oblique. Oblique rotations rotate the PCs in a way that emphasizes the clustering of the variables while allowing scores located on different PCs to be correlated. Orthogonal rotations rotate the PCs while maintaining the orthogonality between scores on different PCs. Varimax rotation is a common technique of orthogonal rotation which maximizes the variance or the sum of squares of the loadings. Ultimately, the amount of variance explained by each factor are changed (usually lower) but the amount of total variance explained remains the same.[12]

When considering how many principal components to retain for further analysis, one has many options. The Kaiser criterion introduced by Kaiser in 1960, simply states that only factors with eigenvalues greater than one are retained.[12] Principal components are the eigenvectors and the corresponding eigenvalues are the variances that are extracted by the PC.[11] However, sometimes this method retains too many PCs. The scree test introduced by Cattell in 1966 depends on the scree plot; a plot of the eigenvalues and the successive number of principal components (shown in Figure 1.4.2). It is common to see a steep decrease in the values of the eigenvalues followed by a slow steady decrease. Therefore, the few eigenvalues that form the steep decrease are retained while the remaining slow steady eigenvalues are “factorial scree” or debris. However, this technique may discard too many PCs.[12] The last technique and the one used is this study involves setting a value of total variance (i.e. 95%) that will be retained. The successive PCs that explain 95% of the total variance are retained and the remaining PCs discarded.

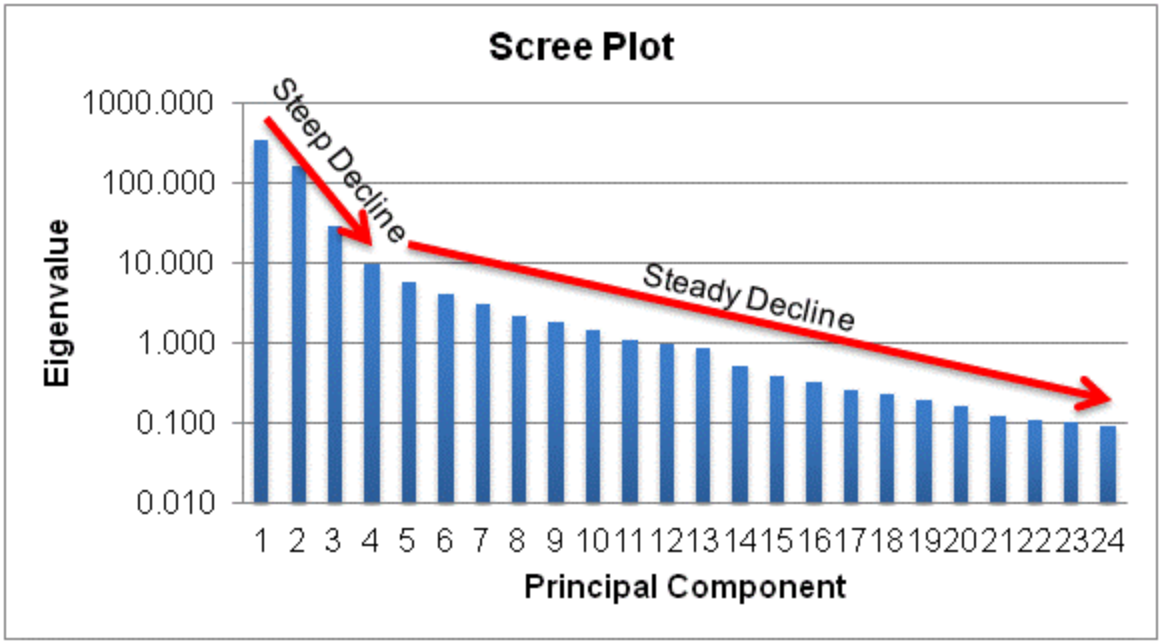


Figure 1.4.2 Example of a PCA scree plot (courtesy of Eric Reichard)

In summary, PCA is the most commonly encountered multivariate statistical technique. It makes the interpretation of large data sets more manageable by creating new axes (PCs) that describe the variables in fewer dimensions. PCA can also identify which variables contribute the most to the total variance for a given PC. PCA has been employed in the analysis of counterfeit coins[14], electrical tape[15,16], gasoline[27-28], gunpowder[29], hair dye[17], headlamp lens material[30], ignitable liquids[31], illicit drugs[18], lubricating oils[20], paints[21], pen ink[22], and soil[24].

1.5. Discriminant Analysis

Discriminant Analysis (DA) is a supervised pattern recognition technique that can also be considered a dimension reduction technique.[1] DA can be used to visualize the data set in two- or three-dimensions, show group characteristics, predict the group membership that a sample belongs to, and identify the factor that discriminate between the predetermined groups. A few assumptions are made when performing DA, the data for the variables of a group should have a

normal distribution, the variance-covariance matrices are equal across all groups, and the means are not correlated with the variables.[12,26] The procedure for DA is comparable to that of PCA with slight differences. PCA maximizes the sum of squares whereas DA maximizes the Fisher ratio which is the ratio of between group variance to within group variance.[1]

First, a new set of axes are created called canonical variates (CVs) or discriminant axes which are linear combinations of the original data as shown in Equation 1.12. The new axes are defined by f_i, while $\bar{x}_{A\ or\ B}$ are the centroids of the classes, and C_{AB} is the pooled variance-covariance matrix for groups A and B. This matrix is calculated according to Equation 1.13 where N_A is the number of samples in group A and N_B is the number of samples in group B. The calculation of C_{AB} is for two groups but this can be expanded if more than two groups are present in the data. The new axes can be plotted against one another to produce an observations plot.

$$f_i = (\bar{x}_A - \bar{x}_B)C_{AB}^{-1}x_{i'} \qquad \text{Equation 1.12}$$

$$C_{AB} = \frac{(N_A-1)C_A+(N_B-1)C_B}{(N_A+N_B-2)} \qquad \text{Equation 1.13}$$

Then, the Mahalanobis distance is calculated for the distance between each sample and the centroid of each group. Equation 1.14 is the Mahalanobis distance calculation where d_{iA} is the distance between sample i and group A, and $\bar{x}_A$ is the centroid of group A, $\bar{x}_B$ is the centroid of group B, C_A is the variance-covariance matrix for group A.

$$d_{iA} = \sqrt{(x_i - \bar{x}_A)C_A^{-1}(x_i - \bar{x}_B)'} \qquad \text{Equation 1.14}$$

Samples are classified into groups based on the smallest distance value. However, in order for this distance calculation to work the number of samples must be greater than the number of variables otherwise the variance-covariance would not have an inverse. PCA is optionally performed before DA to reduce the number of variables.[8]

Once the samples have been classified the accuracy of the classification rule is examined. Cross-validation determines the probability of the sample belonging to each of the groups by removing each sample in turn from the data set, creating a classification rule, and predicting the classification of the sample. There are several ways to carry out validation. Resubstitution uses the entire data set as a training set to develop the classification rule based on the known classification of the samples in the data set. The class membership of each sample is then predicted based on the model and cross-validation determines the accuracy of the rule; however, this method tends to overestimate the classification accuracy. The second method is the hold-out method which separates the data set into a training set and a test set. The training set is used to develop the classification rule while the test set is used to check the classification rule. This method is unbiased but requires a rather large data set to begin with; this method may be inappropriate. The last method that will be discussed is the leave-one-out method which removes each sample from the data set in turn and creates the classification rule based on the remaining samples. The left out sample is then reclassified based on the created classification rule. If desired more than a single sample may be removed at a given time.[1] These methods place the results in a table called the confusion matrix an example of which is shown in Table 1.1.1.

Table 1.5.1 Example of a DA confusion matrix (courtesy of John Goodpaster)

Predicted State / Actual State	IN	MA	MN	MI	MD	Total	% Correct
IN	1	0	0	1	0	2	50.00%
MA	0	2	0	0	0	2	100.00%
MN	0	0	2	0	0	2	100.00%
MI	0	0	0	3	0	3	100.00%
MD	0	0	0	0	2	2	100.00%
Total	1	2	2	4	2	11	90.91%

In summary, DA aims to minimize the within-sample variance and maximize the between-sample variance by creating new axes (CVs). In other words, bring similar samples as close together as possible but make samples not similar to the samples as far away as possible. After the CVs are constructed, the distance between each sample and the centroid of each group is calculated. The sample is then placed into the group with the smallest distance. Once all the samples have been grouped the accuracy is checked by cross-validation. DA has been used for the analysis of electrical tape[15,16], gasoline[32], hair dye[17], headlamp lens material[30], lubricating oils[20], pen ink[22], soil[24], and tire rubber.[33]

1.6. Analysis of Variance

ANOVA tests whether the differences between sample means is larger than what can be explained by random error.[34] ANOVA tests for significant differences between means for the independent groups by partitioning the total variance according to their source. ANOVA is comparable to the t test but without the loss of assurance. For example, performing t tests on five samples repeatedly at the 0.05% confidence level; the probability of making a correct decision on whether the averages differ significantly for the first pair of averages would be 95%. The probability of making a correct decision for the second pair of averages would also be 95% but the probability of making both decisions correctly would be 95% of 95% or approximately 90%. So by the tenth t test the probability of making a correct decision would be reduced to approximately 60%. However, with ANOVA the assurance of making a correct decision is not diminished with larger data sets.[35]

ANOVA relies on the fact that the total variance can be partitioned into the variance between groups and within groups. Remember, variance is the sum of squared deviations from the mean divided by the sample size minus one (the degrees of freedom). The total mean sum of squared deviation is calculated as

shown Equation 1.15 where $\bar{x}$ is the overall average, h is the number of samples, and n is sample size.

$$MSS_{TOTAL} = \sum_i \sum_j (x_{ij} - \bar{x})^2 / hn - 1 \qquad \text{Equation 1.15}$$

ANOVA begins with the null hypothesis that the within-sample and the between-sample variances are equal. The between-sample mean sum of squared deviation is estimated for each sample using Equation 1.16 where n is the number of variables, $\bar{x_i}$ is the average of sample i, $\bar{x}$ is the overall average, and h is the number of samples.

$$MSS_{BETWEEN} = n \sum_i (\bar{x_i} - \bar{x})^2 / h - 1 \qquad \text{Equation 1.16}$$

The within-sample mean sum of squared deviation is estimated for each sample according to Equation 1.17 where $\bar{x_i}$ is the average for sample i, h is the number of samples, and n is sample size.

$$MSS_{WITHIN} = \sum_i \sum_j (x_{ij} - \bar{x_i})^2 / h(n-1) \qquad \text{Equation 1.17}$$

The variance of each sample is calculated by the summation over j divided by $(n-1)$ while the summation of i divided by h produces the average of the sample variances. If the null hypothesis is not rejected than there is no significant difference between the estimates. However, if the null hypothesis is rejected then there is a significant difference in the estimates. This indicates that the between-group variance estimate is greater than the within-group variance estimate.[6,34]

This is achieved by the one-tailed F test, which determines whether there is a statistically significant difference between the groups at a selected confidence level (e.g. 0.05%). The F test calculates the ratio of the sum of squared deviations as shown in Equation 1.18 producing an F-value.

$$F - VALUE = \frac{MSS_{BETWEEN}}{MSS_{WITHIN}} \qquad \text{Equation 1.18}$$

This value is calculated for each variable. If this calculated value is greater than the critical value the null hypothesis is rejected, meaning that the difference between the sample means is significant. Lastly, the higher this value is the more significant the variable is. [6,34]

In summary, ANOVA is an appropriate technique to test for statistically significant differences in variable means. ANOVA divides the total variance into between-group variance and within-group variance. Therefore, ANOVA can identify variables that contribute the most to the discrimination between groups. ANOVA is fairly new to forensic science and has been utilized in the analysis of document paper[36], glass[37], gunshot residue[38], and headlamp lens materials[30].

PART 1: FIBER DYE ANALYSIS

CHAPTER 2. INTRODUCTION

This study was suggested by the Microtrace Unit of the Indiana State Police Forensic Laboratory and supported by the Midwest Forensic Research Center.[39] The aim of this study was to evaluate the reproducibility of fiber dye absorbance spectra collected by UV-Vis microspectrophotometry. This was achieved in two parts. The first part was the evaluation of spectra obtained on an instrument at Indiana University Purdue University Indianapolis (IUPUI). The second part of the study was the evaluation of spectra obtained on an instrument at the Indiana State Police (ISP). The overall, evaluation included four chemometric techniques (AHC, PCA, DA, and ANOVA) and a student t-test to determine if a statistically significant difference existed.

2.1. Introduction to Fibers

Fiber analysis is based on the principle that fibers are physically and chemically differentiable. There are three main groups for fiber type: natural, semisynthetic, and man-made. Natural fibers are those that naturally occur in nature. Semisynthetic fibers have been manufactured from natural fiber forming polymers (regenerated fibers). Manmade fibers are made from synthetic fiber forming polymers.

Within the natural fibers there are three divisions based on the source of the fiber. Natural fibers can be produced from animals (protein), plants (cellulose), or minerals (asbestos). Animal fibers are produced by the animal in the form of fur or hair, which is made of the protein keratin, usually from alpaca, angora rabbit, camel, goat, and sheep (e.g. wool or cashmere). Animal fibers can also be produced from the secreted material, which is made of the protein

fibroin, produced by *Bombyx mori* (mulberry silkworm) from the insect class of *Lepidoptera* (e.g. silk). Plant fibers are divided into three categories depending on what part of the plant the fiber is produced. Plant fibers mainly consist of cellulose, hemi-celluloses, pectins, lignins, water solubles, and fats and waxes. These fibers can be produced from the seed, the leaves, or the stem. Stem fibers are found in bundles in dicotyledonous plants. The mineral fibers come from asbestos.

Man-made fibers can be divided into two subcategories: synthetic fibers and other fibers. Synthetic fibers are those produced from synthetic polymers. Other fibers are those manufactured using other materials such as carbon, glass, or metal.

This discussion will focus on cotton because it is the most commonly seen according to the Indiana State Police forensic laboratory and was used in the study. Cotton is a plant fiber produced from seeds of the species *Gossypium*. The cotton fibers begin to grow when the plant flowers fall off. After two months, the boll (pod) containing the fibers bursts open exposing the fibers and the moisture evaporates. When the moisture evaporates, the fibers cell wall shrinks and twists. These twists allow for greater flexibility during the spinning process. Cotton fibers can be dyed with direct, azoic, vat, sulphur, reactive, and metal complex dyes.[40]

The most common request given to a forensic fiber analyst is to compare fibers to determine whether a common source exists. First, it is determined whether the fiber is natural or synthetic. If natural, the fiber's class is determined (animal, plant, or mineral). If synthetic, the fiber's genus is determined (acrylic, nylon, polyester). Then a testing scheme is derived to try to distinguish the questioned and known fibers.

2.2. Fiber Dyes and Color Theory

A chromophore is a simple unsaturated group that is usually attached to a ring and is responsible for the absorption of UV-vis radiation (e.g. azo, cyan, aldehyde, keto, and carboxylic groups). An auxochrome alters the depth of color (the light absorbed by the chromophore). The auxochrome is a basic salt-forming group; e.g. a hydroxyl or amino group. Together the chromophore and auxochrome create the dye.

It is common to either see the dyes classified according to the method of application, chemical class, or by the type of fiber it can be applied to. It is also common to see many names given to a single dye. The first is the colour index number (CI), a five digit number assigned to all dyes. Dyes of the same class have similar numbers. The second is the generic name which describes the dye class and intended use. The last is the commercial name assigned to the dye by the manufacturer. This may result in several commercial names for a single dye.

When classified according to application there are acid, azoic, basic, direct, dispersive, metallized, reactive, sulphur, and vat dyes.[41] Acid dyes are applied under acidic conditions to polyamide, wool, silk, and polypropylene fibers. Ionic or salt linkage interactions occur between the dye molecule and the fibers polymeric structure.[41] Azoic dyes consist of a napthol and a diazo component that cause a coupling reaction with each other. The napthol is applied to the fibers first; then the diazo which forms an insoluble compound with the napthol.[41] Basic dyes are also applied under acidic conditions and heat but to polyacrylonitrile, modified acrylic, polyester, and polypropylene fibers. These dyes are ammonium, sulphonium, or oxonium salts that ionize in solution so that the color component is in cationic form; which will react with the negatively charged fiber surface to neutralize the charge.[41] Direct dyes are applied in the presence of an electrolyte and heat to a cellulosic fiber. The electrolyte neutralizes the fiber's surface allowing the dye to enter the fiber. Dispersive dyes are applied primarily to polyester and acetate fibers; occasionally applied to polyacrylonitrile, polyamide, and polypropylene fibers. Hydrogen bonding and/or

van de Waals' forces hold the dye to the fiber.[41] Metallized dyes are used on wool and polypropylene fibers. There are three methods of complex formation: chrome mordant method, metachrome method, and after-chrome method. For the chrome mordant method, the mordant is applied to the fiber first and then the dye is applied forming a complex. In the metachrome method, the dye and mordant are applied simultaneously. The after-chrome method involves the application of an acid dye first then the addition of the mordant to form the complex. Other methods exist like incorporating the metal in the structure of the dye and then van de Waals' forces ensure bonding to the fiber.[41] Reactive dyes form covalent bonds with the functional groups of cellulosic and wool fibers. The application is similar to direct dyes but use an alkaline medium for cellulosic fibers and acidic for wool fibers.[41] Sulphur dyes contain sulphur atoms in their structure; therefore, require a reducing agent in an alkaline medium to form soluble form of the dye. Once in the cellulosic fiber oxidation occurs to form the insoluble form.[41] Vat dyes are only applied to cellulosic fibers and are insoluble so a reducing agent is required before penetration of the fiber. Once in the fiber the dye is oxidized to the insoluble form.[41]

Colorimetry is the science of transforming the observed color into mathematical form. This can be achieved in two ways: visual comparison with known standards and instrumental measurement. For the visual comparison color atlases or color appearance systems are used as the standards. The most common are the Munsell Book of Colour, the Ostwald system, and the DIN system. The Munsell concept shows that color is three-dimensional. The hue is oriented as a circle with five principal hues: red, yellow, green, blue, and purple. The center of the circle is the vertical achromatic axis (black to white) measured as lightness. The horizontal axis is the measurement of chroma or the intensity of the color. Instrumental analysis has been established by the Commission Internationale de L'Eclairage to calculate the numerical color value by conducting a study where people were asked to reproduce a shade of color using red, blue, and green lights. The three colors simulate the three color receptors in the

human eye which are red, blue, and green. Based on this study, tristimulus values (X, Y, Z) and chromaticity coordinates (x, y, z) were derived. The chromaticity coordinates are calculated by normalizing the tristimulus values as shown in Equation 2.1.

$$x = \frac{X}{X+Y+Z} \qquad y = \frac{Y}{X+Y+Z} \qquad z = \frac{Z}{X+Y+Z} \qquad \text{Equation 2.1}$$

These values can be calculated with transmittance spectra obtained by UV-vis MSP. However, when the spectra are obtained using absorbance instead of transmission the calculated coordinates are the complementary chromaticity coordinates.[42] Studies involving color coordinates were used on dyed cotton[43,44], dyed wool fibers[45], paint fragments[46,47], and dyed polyester, wool, and polyamide fibers.[48,49,50]

2.3. Fiber Analysis

Fiber analysis relies on microscopy, spectroscopy, chromatography, and mass spectrometry. Microscopy can include stereomicroscopy, polarized light microscopy (PLM), fluorescence microscopy, and scanning electron microscopy coupled to energy dispersive x-ray (SEM/EDX). Spectroscopy can include UV-Vis microspectrophotometry (MSP), Fourier-transform infrared microspectroscopy (FTIR), Raman spectroscopy, and X-ray fluorescence spectroscopy (XRF). Chromatography can include thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), and pyrolysis gas-chromatography (Py-GC). Mass spectrometry can be coupled to HPLC and GC.

Stereomicroscopy is primarily used for the recovery of fibers because of the low magnification, large working distance, and dual light source (reflectance and transmittance). PLM is used to study the optical characteristics of the fibers. Each fiber type will have different optical characteristics allowing for discrimination. A fluorescence microscope is used to study the fluorescent ability

of dyes, optical brighteners, or contaminants.[51] SEM/EDX can identify morphological features and cross-sectional shape; though it is seldom used.[52]

The human eye contains three cone-like nerve endings that allow us to perceive color. However, this varies between individuals and can also vary under different lighting conditions. There is also the issue of metamerism which occurs when two objects dyed with different dyes appears to be the same shade. The use of MSP can partially alleviate these problems. The instrument contains two components: the microscope and the spectrophotometer. Unlike the other techniques, this technique depends solely on the presence of conjugated bonds.[42] Infrared microspectroscopy can be utilized if the analysis via PLM and MSP conclude that the fibers are indistinguishable. This technique allows the analyst to identify and compare the polymeric structure of the fiber in question to the set of known fibers. The comparison of dyes via FTIR microscopy is more complicated because variations in the spectra are subtle. Raman spectroscopy can be considered complementary to FTIR and also identifies the chemical composition. However, unlike FTIR spectroscopy, Raman does not rely on a dipole moment but the bending and stretching of the bonds. Also unlike FTIR, Raman can show the characteristics of the dye the fiber is dyed with.[53] The elemental composition can be examined via XRF techniques, especially for carpet and automotive fibers.[52]

It is necessary to extract the dye before performing either TLC or HPLC/MS. These methods allow the separation of visually similar but chemically different dyes. TLC involves choosing the correct extraction solution and elution solution. When determining the optimal eluent, these factors should be considered: the separation of the bands, the sharpness of the bands, the movement from the origin, components travelling close to or at the solvent front, and the strength of the dye extraction. Therefore, this can be a difficult and long process due to the many types of fibers and dye classes.[41] HPLC has some advantages over MSP and TLC for dye analysis. Advantage include: better resolution of the dyes, quantitation if desired, and better reproducibility. However,

like with TLC there are several variables that are dependent on knowledge of the type of dye and the type of fiber. Once these factors are known a method can be developed, which means choosing the right column, mobile phase, and detector. CE may offer greater separation and allows for a greater analyte range.[54] Lastly, Py-GC/MS can be utilized to determine the composition of the fiber based on the decomposition fragments.[52]

CHAPTER 3. MATERIALS AND METHODS

3.1. Materials

Testfabrics, Inc. (West Pittson, PA) provided six red dyed cotton exemplars for analysis. An additional six red dyed cotton exemplars were provided by Dr. Stephen Morgan at the University of South Carolina. These twelve exemplars were dyed with different shades of red. Table 3.1.1 and Table 3.1.2 show images taken via a MSP, lists the names of the dyes, and the label that was assigned by the analyst for easy reference. STC EMPA is a catalog designated for "soil test cloth -EMPA-" and EMPA is a Swiss Federal Testing facility in St. Gall, Switzerland.

Table 3.1.1 Dyed cotton exemplars provided by Testfabrics, Inc.

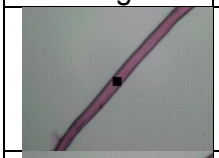
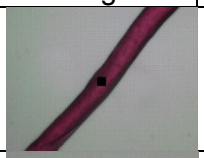
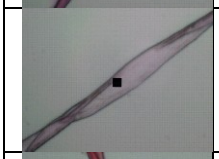
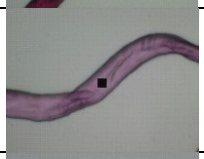
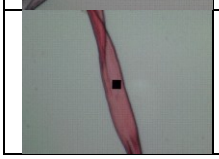
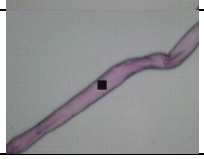
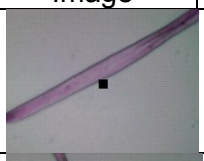
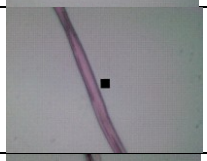
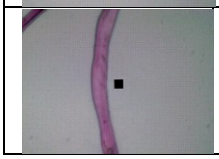
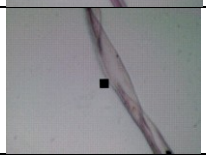
Image	CI Name	Label	Image	CI Name	Label
	Style 419, Bleached Mercerized Cotton Dyed With Direct Red C-380	A		STC EMPA 475 Dyed With Reactive Red 195	D
	Style 460, Bleached Cotton Dyed With Reactive Red 120	B		STC EMPA 495 Dyed With Reactive Red 2	E
	Style 460, Bleached Cotton Dyed With Reactive Red 123	C		STC EMPA 498 Dyed With Reactive Red 228	F

Table 3.1.2 Dyed cotton exemplars provided by Dr. Stephen Morgan

Image	Name	Label	Image	Name	Label
	Direct Red 84	G		Reactive Red 239/241	J
	Reactive Red 180	H		Vat Red 10	K
	Reactive Red 198	I		Vat Red 15	L

3.2. Instrumental Analysis

The twelve exemplars were prepared for analysis following standard MSP protocols outlined by the Scientific Working Group on Materials Analysis (SWGMAT).[55] Ten fibers were removed from each exemplar and mounted in glycerin on glass microscope slides. Glycerin was deemed an appropriate mounting medium according to SWGMAT. However, glycerin is not a good mounting media for the long term storage of materials. At Indiana University Purdue University Indianapolis (IUPUI) a CRAIC QDI 2000 MSP (CRAIC Technologies, San Dimas CA) was used. At the Indiana State Police (ISP) laboratory, a CRAIC QDI 1000 MSP was used. Both MSPs were in transmitted light mode, at 35x magnification, and were calibrated before each use with NIST traceable standards. Autoset optimization, a dark scan, and a reference scan were performed before each set of sample scans. Ten scans were taken, from different locations, for each fiber as absorbance values. Two additional fibers from each dye were mounted and analyzed under the same conditions as the previous samples for external validation. Three different data sets were studied in this research. First, only the exemplars provided by Testfabrics, Inc. were analyzed including external validation samples. Second, all twelve exemplars

were analyzed including external validation samples at IUPUI. Lastly, all twelve exemplars were analyzed at ISP to conduct an inter-laboratory study; no external validation was collected.

3.3. Data Analysis

The three data sets were truncated to a wavelength range of 350-800 nm, the range of visible light. Then, background subtraction was performed by subtracting the minimum absorbance for each sample. The data was then normalized by dividing each absorbance value by the square root of the sum of squares of all absorbance values. This eliminated variability due to fiber thickness and dye uptake along the fiber. Statistical evaluation was performed with Microsoft Excel XLSTAT2010 (Paris, France), an add-in for Excel. For AHC, the ten normalized scans for each fiber were averaged together. For PCA and ANOVA, all scans were utilized instead of the averages. The number of PCs used for DA was determined based on a 95% set limit for the cumulative percent variance. Therefore, for the data set with only the Testfabrics, Inc. exemplars' six PCs were utilized. This resulted in the use of eight and five PCs for the data sets collected at IUPUI and ISP respectively for all twelve exemplars.

The first derivative of each spectrum was calculated with the MSP software for the IUPUI data set with data collected for all twelve exemplars. Statistical techniques were then performed which included DA. All scans were used for analysis.

Color coordinate analysis was performed on the data set collected at IUPUI for all the exemplars. The color coordinates were calculated based on the transmittance data collected with the MSP GRAMS AI Software with Illuminant C and standard observer angle of 2^{o}. Statistical techniques were then performed which included AHC, PCA, and DA on the original data. For AHC the values for the average of the ten scans per fiber were utilized, while all scans were used for PCA and DA. DA was performed on the original data since the number of samples was greater than the number of variables.

CHAPTER 4. RESULTS AND DISCUSSIONS

4.1. Testfabrics, Inc. Exemplars

4.1.1. Training Set

There were slight visible differences between the analyzed dyes (see Table 3.1.1), Direct Red C-380 (A) was a deep shade of red, Reactive Red 123 (C) was a brighter red, and Reactive Red 195 (D) was a darker shade of pink. Reactive Red 120 (B), Reactive Red 2 (E), and Reactive Red 228 (F) were visually similar. Figure 4.1.1 shows the ten fiber average absorbance spectra for each dye. One noticeable difference is the behavior of the maximum wavelength. Some dyes like A, C, and F had a single maximum wavelength. Dyes B, D, and E had two maxima wavelengths in their spectra. Another noticeable difference is the shifted maximum wavelength of Dye C.

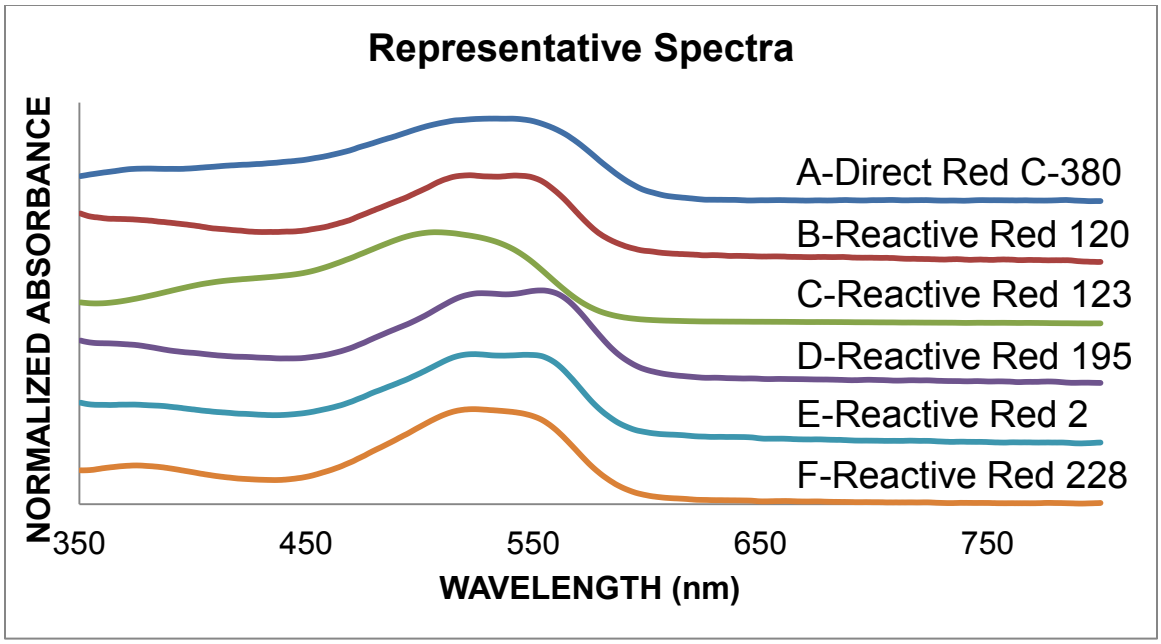


Figure 4.1.1 Representative spectra for Testfabrics, Inc. exemplars

A dendrogram can help detect outliers in the data set. When AHC and PCA were first performed, it was noticed that the ten replicates for Dye B were the most dissimilar. Because of this, the original spectra were examined for possible explanations. The averaged spectra for Dye B showed that Fiber B8 was slightly different than the other B fibers. Upon review of the un-averaged data, it was determined that scan 7 from Fiber B5 as well as scans 2 and 6 from Fiber B8 were outliers. These scans were removed from the data set and all statistical techniques re-run. Ultimately, these scans were removed based on the results from AHC, PCA, and visual inspection of the original normalized data; which showed that these scans were unlike the others from the same fiber. The spectra with the suspicious scans are shown in Figure 4.1.2 with the suspicious scans presented in red.

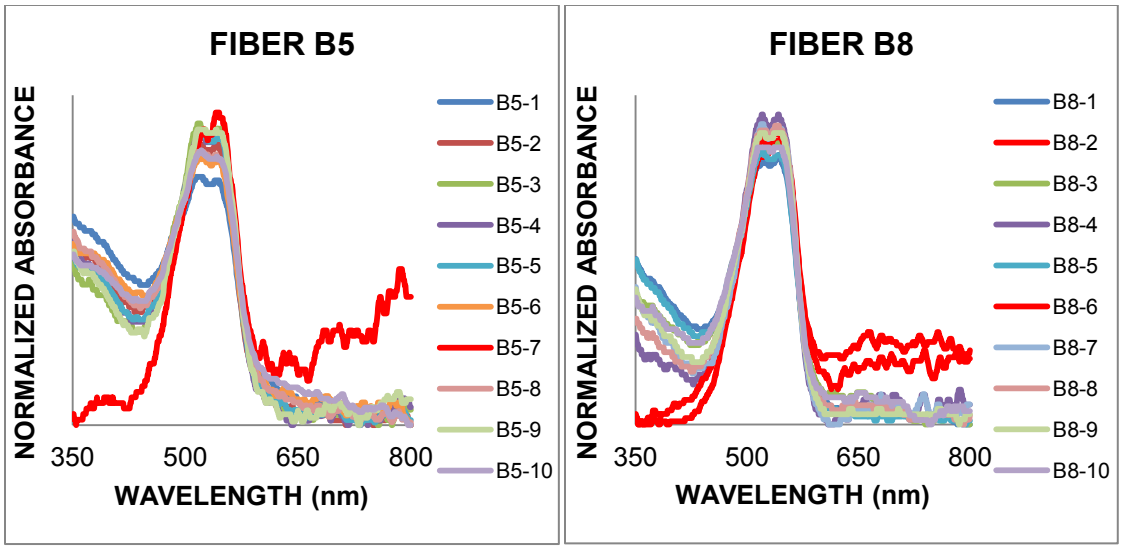


Figure 4.1.2 Detection of outliers for Testfabrics, Inc. exemplars

The AHC dendrogram for fibers A-F minus the outliers is shown in Figure 4.1.3, with four distinct classes being formed based on the truncation line. Bifurcations to the right of this truncation line are more significant when determining the number of classes. Fibers A, C, D, and F clustered well when the dissimilarity was considered. These fibers have low levels of dissimilarity (A at 0.026, C at 0.014, D at 0.014, and F at 0.0025) proving that the replicates are more similar to one another rather than the replicates of other fibers. Also, fibers C and D were clustered so well that they were placed into separate classes. Fibers B and E did not cluster as well and were split between two classes. A Central Objects Plot represents those spectra that are most similar to the average spectra for that class. So, in Figure 4.1.4 the spectra for fibers A7, F5, C7, and D5 represent the average of all spectra for the class they belong to, Class 1, 2, 3, and 4 respectively. As this Figure shows each class has different spectral features. Class 1 has a lower absorbance and a single maximum wavelength. Class 2 has a relatively higher absorbance and a single maximum wavelength. The single maximum wavelength for Class 3 is shifted to the left. Class 4 is the only class to contain two distinct maxima.

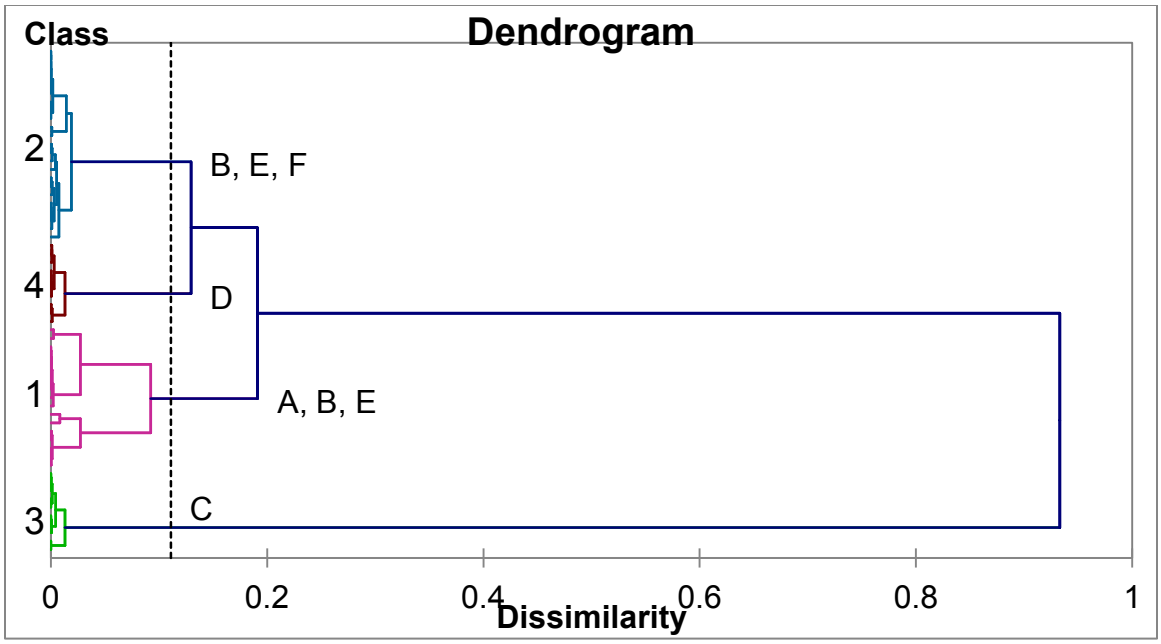


Figure 4.1.3 AHC dendrogram for Testfabrics, Inc. exemplars

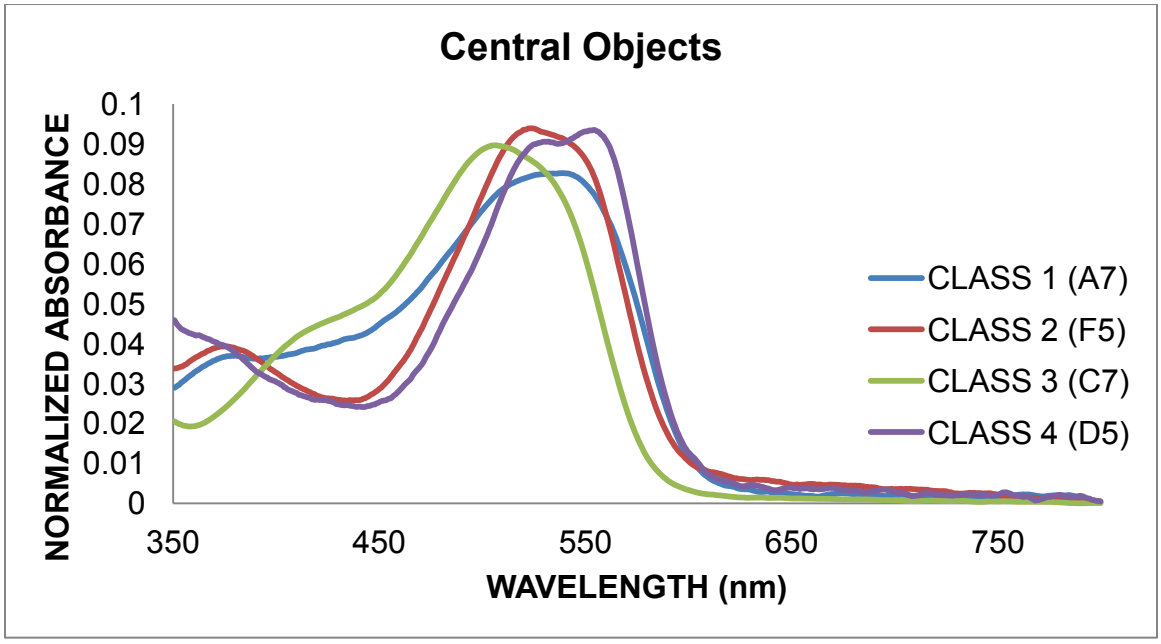


Figure 4.1.4 AHC central objects plot for Testfabrics, Inc. exemplars

For PCA every scan was used instead of using the averages. The observations plot from PCA showed that the first two principal components captured 67.51% of the total variance (data not shown). From that observations plot, Class 3 which contained Dye C was distinguishable from the other dyes. Dye A was slightly distinguishable from the other dyes. The remaining dyes were not as distinguishable as seen by the overlap. To obtain a better view of the dimension of the data, PCA was performed utilizing Varimax rotation by 3 factors. The total variance captured decreased to 56.87%. The observations plot in Figure 4.1.5 is coded according to dye. Now, Dyes A, C, and D were clearly distinguished from the other dyes. Dyes B, E, and F are still overlaid which indicated it is harder to distinguish between these dyes.

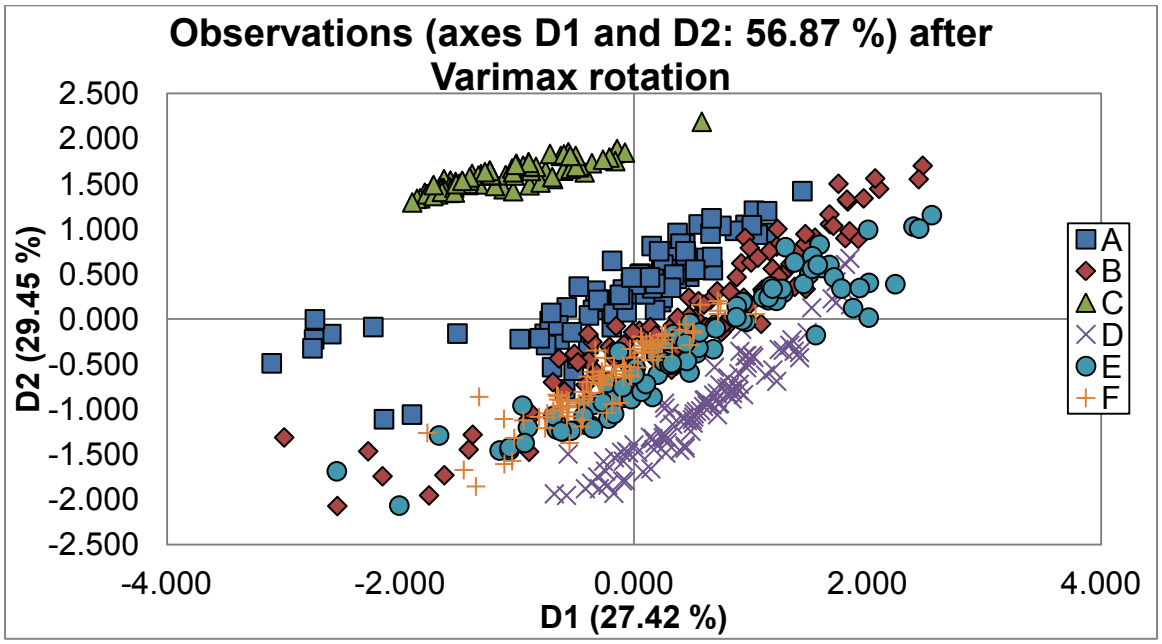


Figure 4.1.5 PCA observations plot for Testfabrics, Inc. exemplars

A factor loadings plot of the original variables can be used to illustrate the PCA coordinate system. However, the factor loadings plot for this data set is slightly different in that the factor loadings are in the form of the cosine of the angle between the principal component and the variable. Where the cosines are positive correspond to positive correlations. Negative cosines are areas of

negative correlations. If the cosine is close to zero there is no correlation. Therefore, the main area of positive correlation for the first principal component occurred between 350-387 nm and the main area of negative correlation occurred between 502-513 nm. The main area of positive correlation for the second principal component occurred between 402-484 nm, while the main area of negative correlation occurred between 528-565 nm. These regions were overlaid with the representative spectra to produce Figure 4.1.6. The areas of correlation between the first principal component and the wavelengths are shaded light blue; the areas of correlation between the second principal component and the wavelengths are shaded light red. Therefore, the first principal component exhibits a strong positive correlation with the absorbance in the higher wavelength region and a strong negative correlation to the region that contains the maximum wavelength of Dye C. The second principal component exhibits a strong positive correlation to the region just left of the maxima wavelengths and a strong negative correlation to the region containing the maxima wavelengths of the remaining fibers.

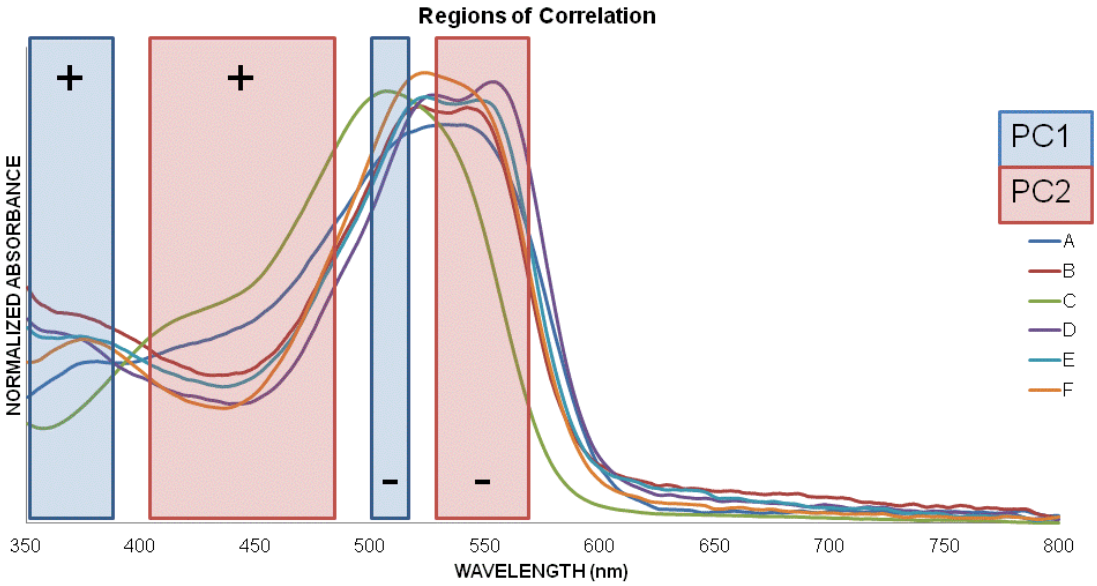


Figure 4.1.6 PCA regions of correlation for Testfabrics, Inc. exemplars

DA was performed several times using the rotated factor scores. At first, the fibers were assigned to one of four groups which were chosen based on the AHC classes. This resulted in an overall classification accuracy of 90.45% (data not shown). The most confusion occurred between Classes 1 and 2. Classes 3 and 4 were correctly classified 100% of the time. The second DA assigned each fiber to one of six classes; according to dye. The result of this is shown as the observations plot in Figure 4.1.7. The classification accuracy was 85.26%. This showed that the most confusion occurred between Reactive Red 120, Reactive Red 2, and Reactive Red 228. Direct Red C-380 and Reactive Red 123 were correctly classified 100% of the time. Reactive Red 195 was classified correctly 98% of the time. The confusion matrix can be seen in Table 4.1.1. Lastly, each fiber was assigned to one of sixty groups according to fiber (data not shown). The overall classification accuracy was 35.01% (data not shown). To simplify the confusion matrix and examine what's being misclassified as another dye, a new confusion matrix was created (data not shown). If the fiber was classified as the correct dye it was considered correct. Example: if fiber A1 was classified as A1, A2, •••, or A10 it was considered to be correct but if it was classified as B1, B2, •••, or B10 it was incorrect. The classification accuracy was 91.67%. Overall, fibers from Dye A were misclassified as other fibers from Dye A. Same was seen for fibers from Dyes C and D. Ultimately, these fibers were not unique to themselves but to the general dye. The most confusion was seen between fibers from Dye B which were misclassified as belonging to Dyes E and F and vice versa. The same trend was seen through all three DA results; Dyes A, C, and D were unique while there was confusion between Dyes B, E, and F.

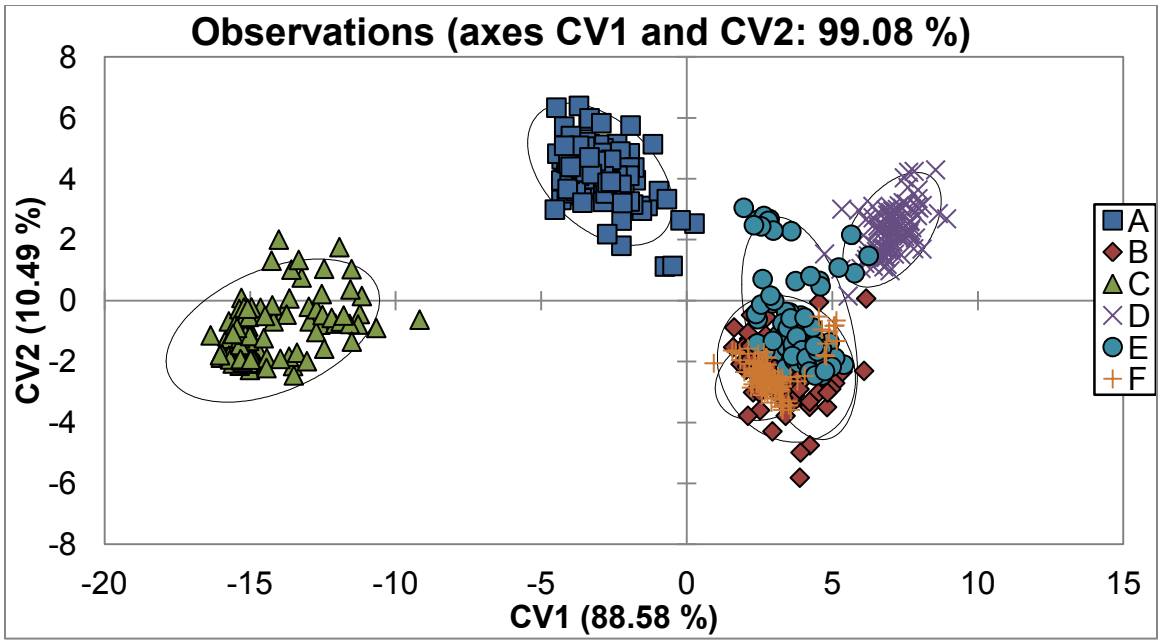


Figure 4.1.7 DA observations plot for Testfabrics, Inc. exemplars

Table 4.1.1 DA confusion matrix for Testfabrics, Inc. exemplars

Predicted Dye Actual Dye	A	B	C	D	E	F	Total	% Correct
A	100	0	0	0	0	0	100	100.00%
B	0	67	0	1	7	22	97	69.07%
C	0	0	100	0	0	0	100	100.00%
D	0	0	0	98	2	0	100	98.00%
E	0	20	0	2	59	19	100	59.00%
F	0	5	0	0	10	85	100	85.00%
Total	100	92	100	101	78	126	597	85.26%

Finally, although PCA helped find the areas of the spectra where there was the most variance, ANOVA helps find the areas where the most discrimination occurs. These areas were overlaid with the representative spectra to produce Figure 4.1.8. The regions of most discrimination occur between 483-492 nm and 566-576 nm. Therefore, the most discrimination occurred on either side of but did not include the maximum wavelengths.

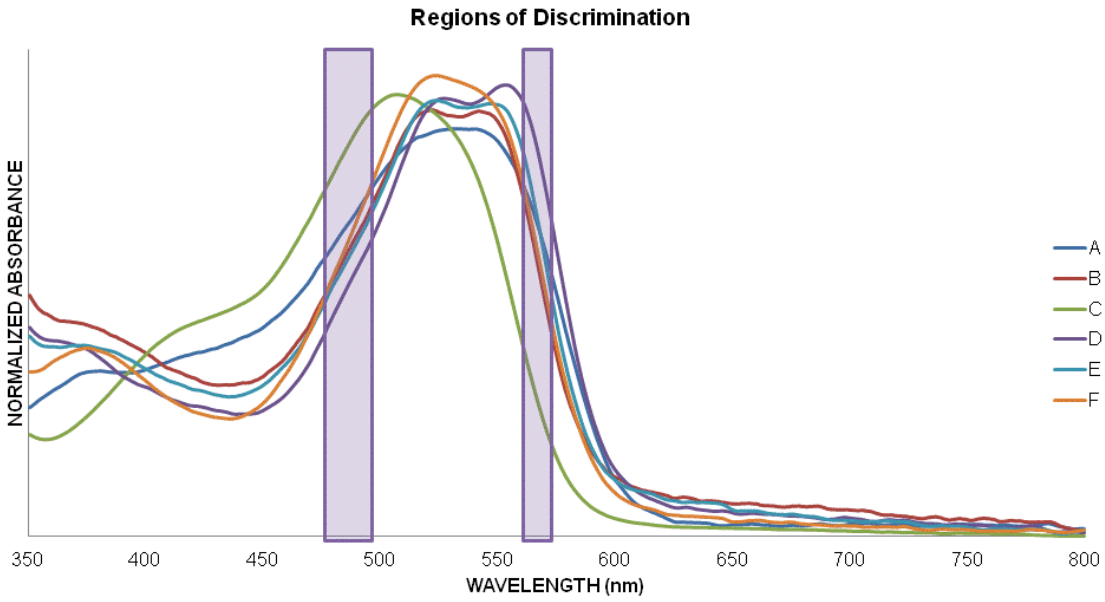


Figure 4.1.8 ANOVA regions of discrimination for Testfabrics, Inc. exemplars

4.1.2. External Validation

DA was performed including the external validation samples to predict which of the six groups the fibers belonged to. The results for the six class prediction can be seen in Figure 4.1.9. This resulted in an overall prediction accuracy of 82.50%. As seen by the confusion matrix in Table 4.1.2. Most of the confusion occurred between the fibers from Dyes B, E, and F. This is consistent with the training set. All of the fibers from Dyes A and C were predicted correctly.

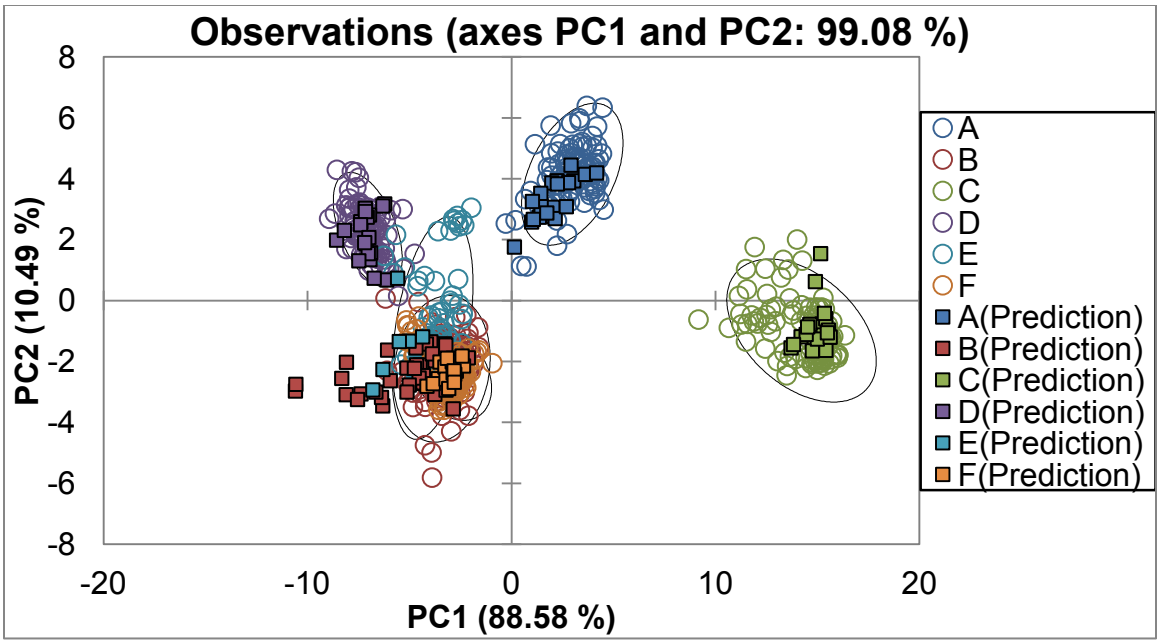


Figure 4.1.9 EV DA observations plot for Testfabrics, Inc. exemplars

Table 4.1.2 EV DA confusion matrix for Testfabrics, Inc. exemplars

Predicted Dye Actual Dye	A	B	C	D	E	F	Total	% Correct
A	20	0	0	0	0	0	20	100.00%
B	0	18	0	0	1	1	20	90.00%
C	0	0	20	0	0	0	20	100.00%
D	0	0	0	19	1	0	20	95.00%
E	0	16	0	0	4	0	20	20.00%
F	0	2	0	0	0	18	20	90.00%
Total	20	36	20	19	6	19	120	82.50%

4.2. All Dyed Exemplars Analyzed at IUPUI

4.2.1. Training Set

To begin analysis, pictures were taken with the UV-Vis MSP (see Tables 3.1.1 and 3.1.2). Based on these pictures there are slight visible differences between the analyzed dyes, Direct Red C-380 (A) was a deep shade of red, Reactive Red 123 (C) was a brighter red, Reactive Red 195 (D) was a darker shade of pink, and Direct Red 84 (G) was red-orange. Reactive Red 120 (B), Reactive Red 2 (E), Reactive Red 228 (F), Reactive Red 180 (H), Reactive Red 198 (I), Reactive Red 239/241 (J), Vat Red 10 (K), and Vat Red 15 (L) are visually similar. Figure 4.2.1A shows the ten-fiber average absorbance spectra for each exemplar provided by Testfabrics, Inc. One noticeable difference is the behavior of the maximum wavelength. Dyes A, C, and F have a single maximum wavelength. Dyes B, D, and E have two maxima in their spectra. Another noticeable difference is the slightly shifted maximum wavelength of Dye C. Figure 4.2.1B shows the ten-fiber average absorbance spectra for each exemplar provided by Dr. Stephen Morgan. Again, there are two distinct behaviors of the maximum wavelength. Dyes G, I, and K have a single maximum wavelength. Dyes H, J, and L have two maxima wavelengths in their spectra. Dye G also had a shifted maximum wavelength. The last noticeable difference is the low absorbance and slightly shifted maximum wavelength of Dye L.

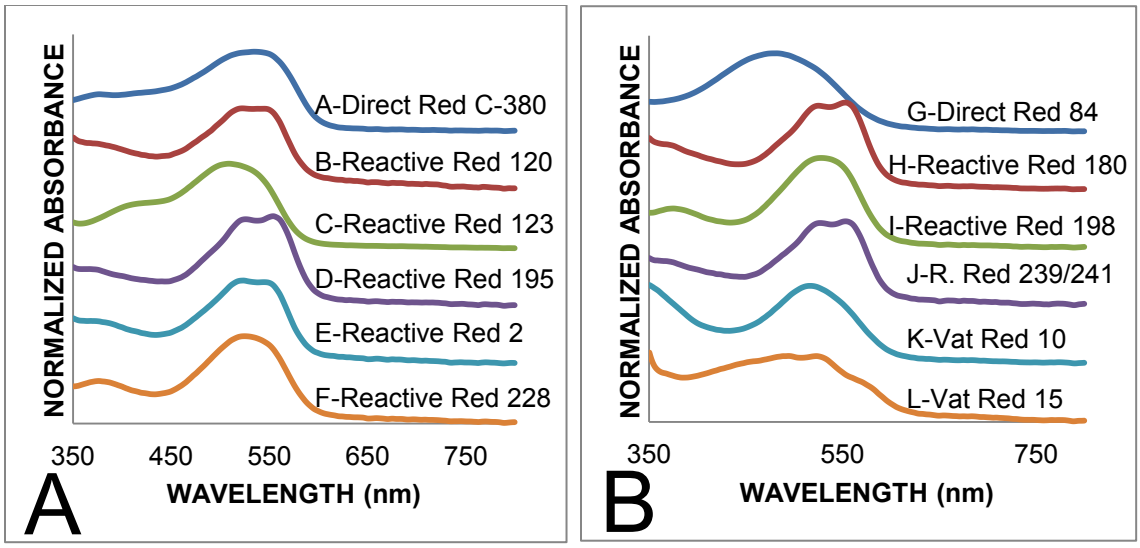


Figure 4.2.1 Representative spectra for all exemplars analyzed at IUPUI

The AHC dendrogram for fibers A-L is shown in Figure 4.2.2, with three distinct classes being formed based on the automatic truncation line. Bifurcations to the right of this truncation line are more significant when determining the number of classes. Looking more closely at the dendrogram, the replicates from Dyes A, C, G, K and L clustered well. These fibers have low levels of dissimilarity (A at 0.0023, C at 0.0043, G at 0.0034, K at 0.0043 and L at 0.0030) proving that the replicates are similar. The replicates from Dyes B, D, E, F, H, I, and J did not cluster well and were split within their respective classes. The replicates for Dye B were clustered with replicates from Dyes E, F, and I indicated by the top star on the dendrogram. However, the replicates from Dye I were mostly well clustered except for one replicate. Dye D was clustered with Dyes H and J indicated by the bottom star on the dendrogram. A central objects plot (Figure 4.3.3) illustrates those fibers that resemble the average spectra representative of the class; ultimately giving the class characteristics. Class 1 contained two maxima wavelengths while the others contained a single maximum wavelength. Classes 2 and 3 had a slightly shifted maximum wavelength. Each Class decreases in total absorbance, where Class 1 had the highest and Class 3 the lowest.

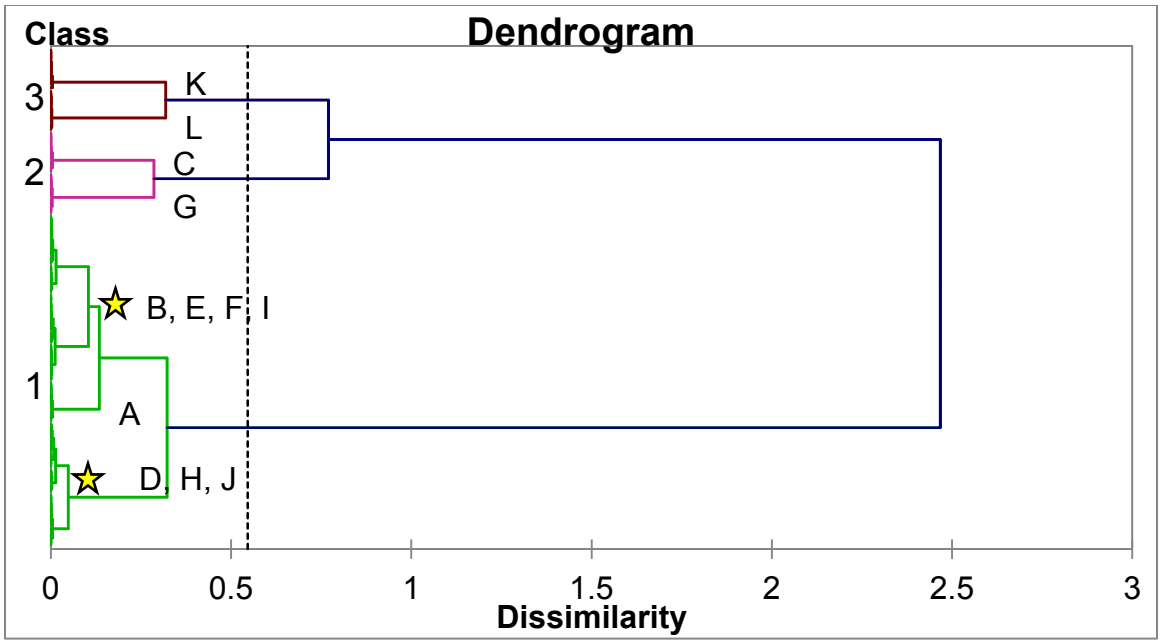


Figure 4.2.2 AHC dendrogram for all exemplars analyzed at IUPUI

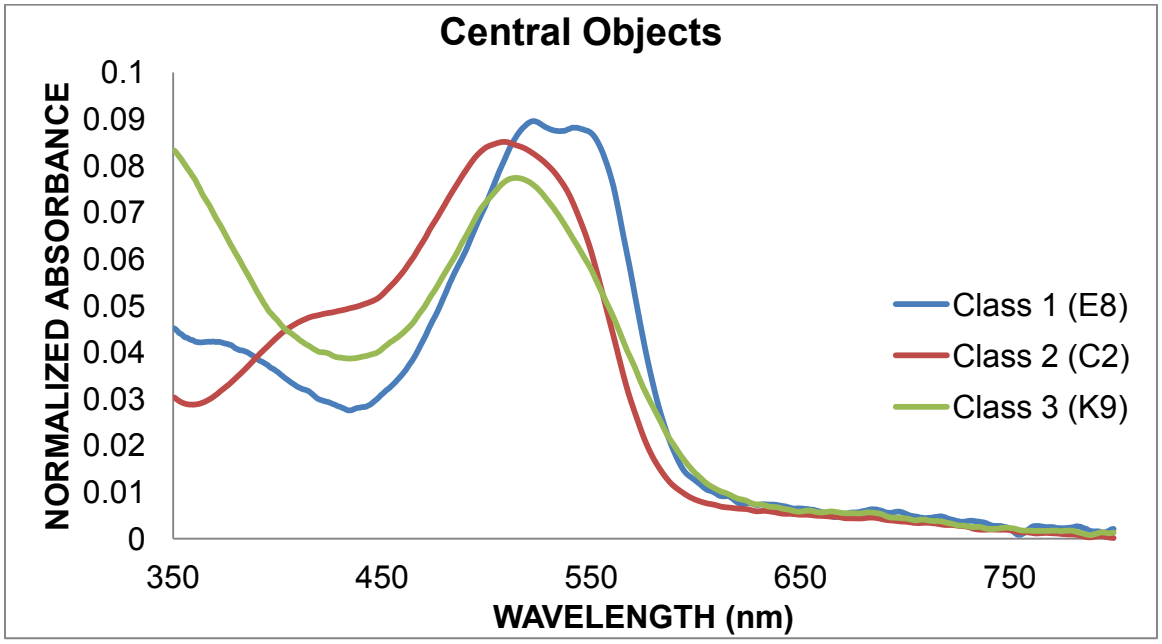
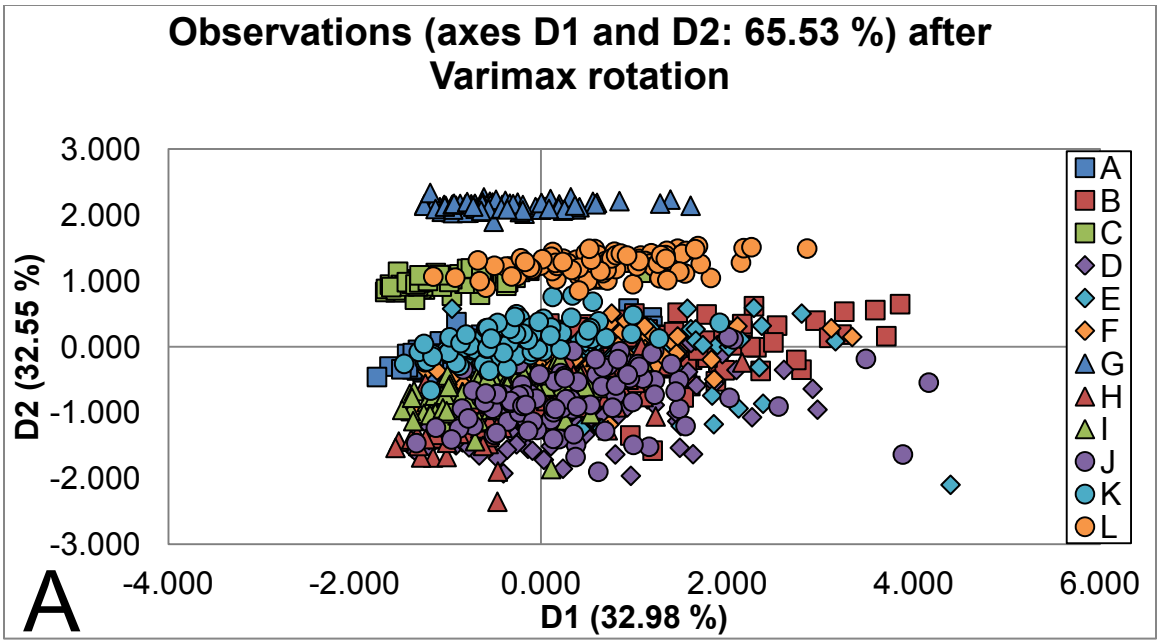


Figure 4.2.3 AHC central objects plot for all exemplars analyzed at IUPUI

The observations plot from PCA showed that the first two principal components captured 68.93% of the total variance (data not shown). Class 2 which contained Dyes C and G were readily distinguishable from the other dyes and each other. Dye L from Class 3 was readily distinguishable while Dye K was not distinguishable. The remaining dyes are not as distinguishable. To obtain a better view of the dimension of the data, PCA was performed utilizing Varimax rotation by 3 factors; the results shown in Figure 4.2.4. Figure 4.2.4A shows the first and second principal components plotted against one another, while Figure 4.2.4B shows the second and third principal components plotted against one another. For easy visualization the data is coded according to dye. The total variance captured decreased to 65.53%. Now, it is more apparent that Dyes C, G, and L were clearly distinguished from the other dyes. There is slight separation of Dyes A and K from the remaining dyes but not between each other. Dyes B, E, F, H, I, and J are overlapping which indicating it is harder to distinguish between these dyes; although not much overlap was seen between the two groups in Class 1.



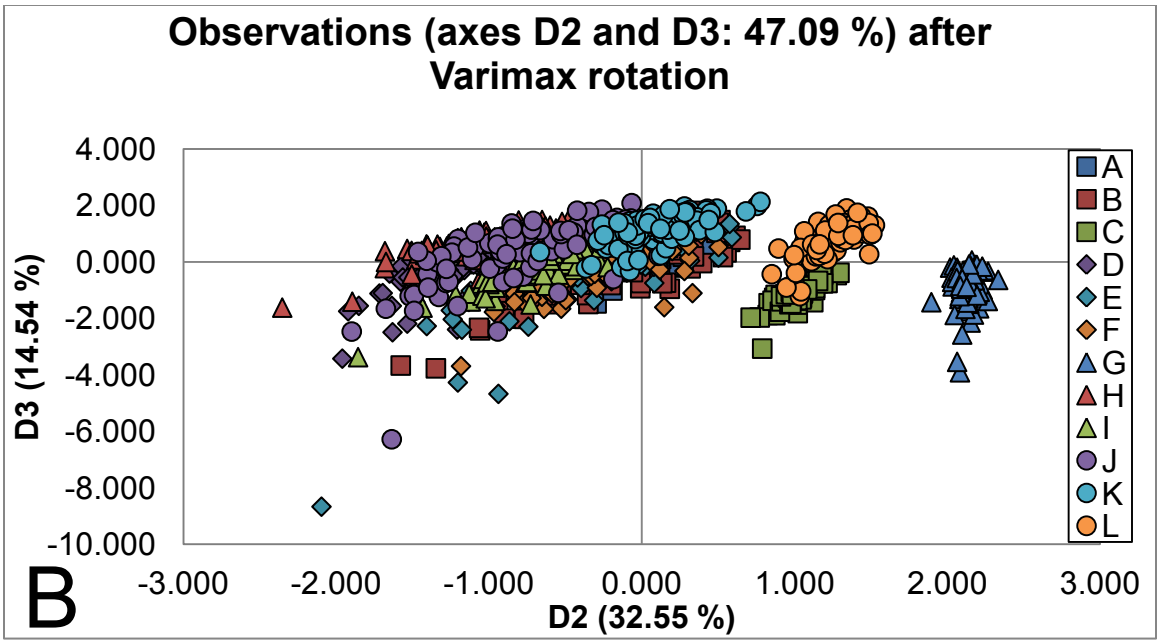


Figure 4.2.4 PCA observations plot for all exemplars analyzed at IUPUI

A factor loadings plot of the original variables can be used to illustrate the PCA coordinate system. The factor loadings are in the form of the cosine of the angle between the principal component and the variable. If the cosines are positive a positive correlation occurs between the variables. When the cosines are negative there is a negative correlation between the variables. A cosine close to zero indicates that there is no correlation between the variables. Figure 4.2.5 shows the correlation regions overlaid with the representative spectra; where the correlation between the first principal component and the wavelengths is shaded light blue while the correlation between the second principal component and the wavelengths is shaded light red. Therefore, the main area of negative correlation for the first principal component occurred between 494-511 nm; the main area of positive correlation occurred between 603-751 nm. The negative correlation range included the maximum wavelength of Dye C. The positive correlation range included the tails after the maximum wavelengths. The main area of positive correlation for the second principal component occurred between 403-491 nm and the main area of negative correlation occurred between 517-577 nm. The positive correlation range included the maximum wavelength of Dye G and rise before the maximum wavelengths of the other dyes. The negative correlation range separated the single maximum wavelength dyes from the double maxima dyes.

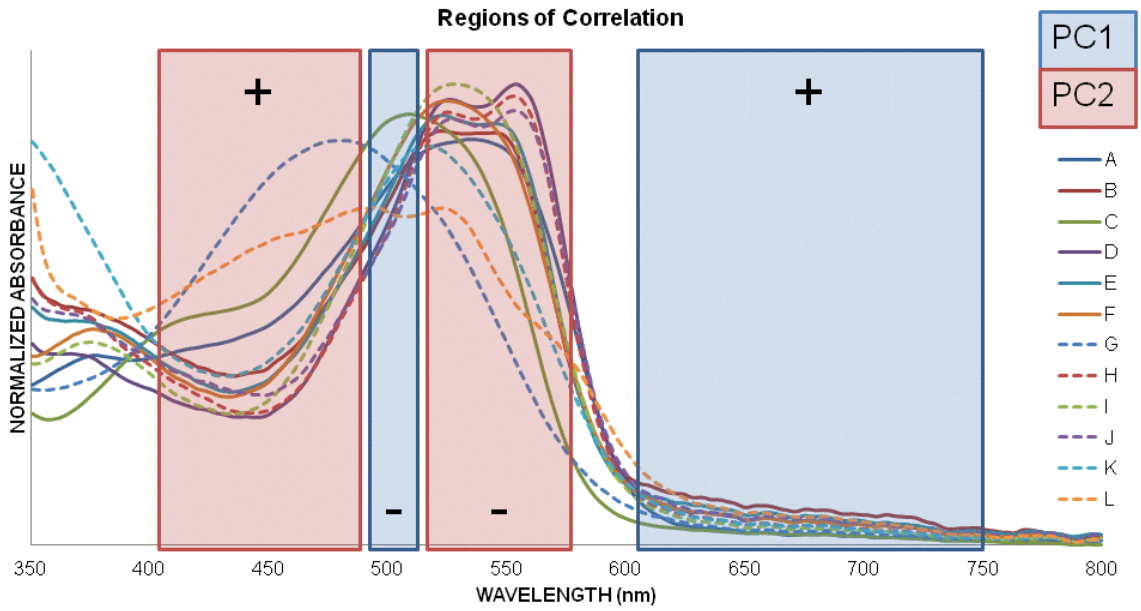


Figure 4.2.5 PCA regions of correlation for all exemplars analyzed at IUPUI

DA was performed several times. At first, the fibers were assigned to one of three groups which were chosen based on the AHC classes. This resulted in an overall classification accuracy of 100.00% (data not shown). Each fiber was correctly placed into one of the three classes. The second DA assigned each fiber to one of twelve classes, one class for each of the dyes. The observations plot is shown in Figure 4.2.6. This figure is coded according to dye. The classification accuracy was 85.00%. This showed that the most confusion occurred between Dyes B, E, F, and I. There was also confusion between Dyes D, H, and J. Dyes A, C, G, K, and L were correctly classified 100% of the time. The confusion matrix is shown in Table 4.2.1. Lastly, each fiber was assigned to one of one hundred and twenty groups according to fiber. The overall classification accuracy was 22.50% (data not shown). To put this in perspective, the classification accuracy was calculated manually (data not shown). If the fiber was classified as the correct dye it was considered correct. Example: if fiber A1 was classified as A1, A2, •••, or A10 it was considered to be correct but if it was classified as B1, B2, •••, or B10 it was incorrect. The classification accuracy was 87.17%. The fibers from Dyes A, C, G, K, and L were only misclassified as those

from the same dye. Fibers from Dye B were misclassified as fibers from Dye E and vice versa. However, fibers from Dye E were also misclassification between as fibers from Dye I. Dye D was misclassified mostly as fibers from Dye J. Some fibers from Dye F were misclassified as fibers from Dyes B, E, and I but 91.00% of the time they were classified as fibers within Dye F. Fibers from Dye H were misclassified as Dyes D and J but 91.00% of the time was classified as fibers within Dye H. Lastly, fibers from Dye I were misclassified as fibers from Dyes E and F but 92.00% of the time were classified as fibers within Dye I.

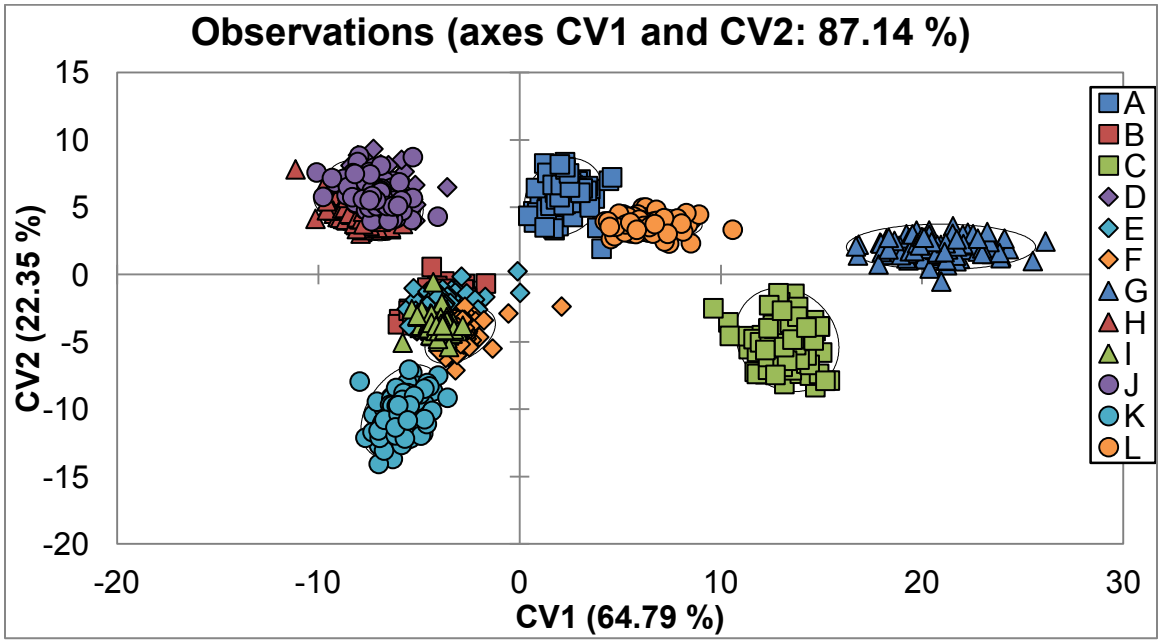


Figure 4.2.6 DA observations plot for all exemplars analyzed at IUPUI

Table 4.2.1 DA confusion matrix for all exemplars analyzed at IUPUI

Predicted Dye / Actual Dye	A	B	C	D	E	F	G	H	I	J	K	L	Total	% Correct
A	100	0	0	0	0	0	0	0	0	0	0	0	100	100.00%
B	0	67	0	0	31	0	0	0	2	0	0	0	100	67.00%
C	0	0	100	0	0	0	0	0	0	0	0	0	100	100.00%
D	0	0	0	72	0	0	0	1	0	27	0	0	100	72.00%
E	0	39	0	0	50	1	0	0	10	0	0	0	100	50.00%
F	0	9	0	0	0	83	0	0	8	0	0	0	100	83.00%
G	0	0	0	0	0	0	100	0	0	0	0	0	100	100.00%
H	0	0	0	5	0	0	0	89	0	6	0	0	100	89.00%
I	0	1	0	0	2	5	0	0	92	0	0	0	100	92.00%
J	0	0	0	22	0	0	0	11	0	67	0	0	100	67.00%
K	0	0	0	0	0	0	0	0	0	0	100	0	100	100.00%
L	0	0	0	0	0	0	0	0	0	0	0	100	100	100.00%
Total	100	116	100	99	83	89	100	101	112	100	100	100	1200	85.00%

A=Direct Red C-380 B=Reactive Red 120 C=Reactive Red 123 D=Reactive Red 195	E=Reactive Red 2 F=Reactive Red 228 G=Direct Red 84 H=Reactive Red 180	I=Reactive Red 198 J=Reactive Red 239/241 K=Vat Red 10 L=Vat Red 15

Finally, ANOVA was performed to find the areas of the spectra where the most discrimination occurs. A plot of the F-values as a function of wavelength (data not shown) reveals which variables aided the discrimination between samples. The higher the F-value the more discrimination achieved at that variable. Therefore, the regions of discrimination occurred between 473-488 nm and 563-575 nm. These regions were overlaid with the representative spectra to produce Figure 4.2.7. This showed that the most discrimination occurred by the maximum wavelength of Direct Red 84 and on the decreasing side of the other maximum wavelengths.

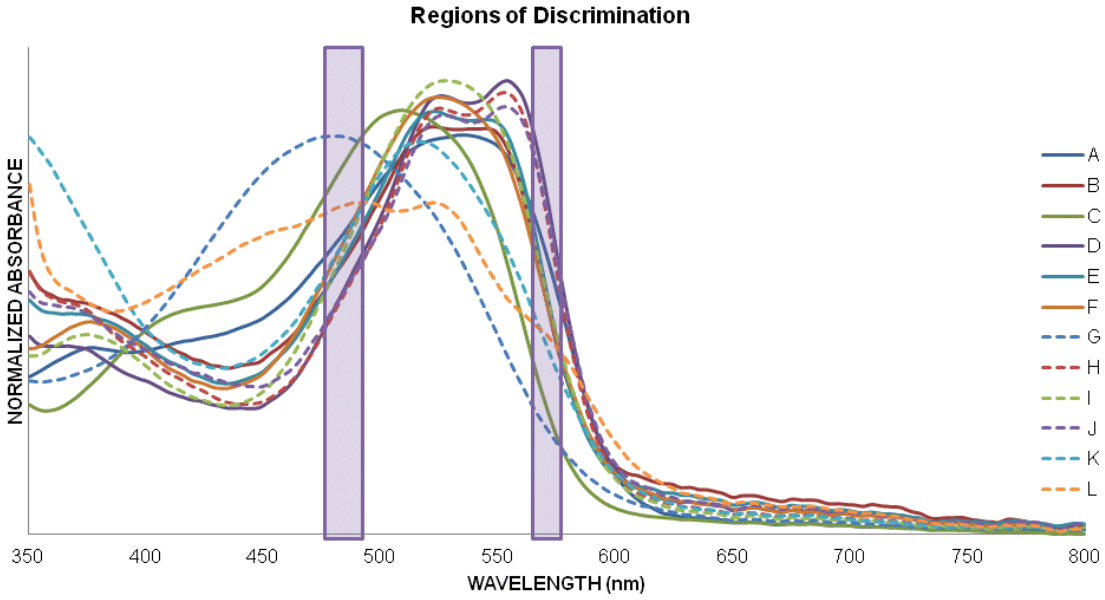


Figure 4.2.7 ANOVA regions of discrimination for all exemplars analyzed at IUPUI

4.2.2. External Validation

DA was performed again this time including the external validation samples. DA was performed to predict which of the twelve groups the twenty four fibers belonged to. This resulted in an overall prediction accuracy of 80.83%. As shown in Table 4.2.2, confusion occurred between the fibers from Dyes B and E; Dyes D, H, and J; and Dyes F and I. All of the fibers from Dyes A, C, G, K, and L were predicted correctly.

Table 4.2.2 EV DA confusion matrix for all exemplars analyzed at IUPUI

Predicted Dye / Actual Dye	A	B	C	D	E	F	G	H	I	J	K	L	Total	% Correct
A	20	0	0	0	0	0	0	0	0	0	0	0	20	100.00%
B	0	10	0	0	10	0	0	0	0	0	0	0	20	50.00%
C	0	0	20	0	0	0	0	0	0	0	0	0	20	100.00%
D	0	0	0	19	0	0	0	0	0	1	0	0	20	95.00%
E	0	5	0	0	15	0	0	0	0	0	0	0	20	75.00%
F	0	0	0	0	0	15	0	0	5	0	0	0	20	75.00%
G	0	0	0	0	0	0	20	0	0	0	0	0	20	100.00%
H	0	0	0	2	0	0	0	16	0	2	0	0	20	80.00%
I	0	1	0	0	1	8	0	0	10	0	0	0	20	50.00%
J	0	0	0	2	0	0	0	9	0	9	0	0	20	45.00%
K	0	0	0	0	0	0	0	0	0	0	20	0	20	100.00%
L	0	0	0	0	0	0	0	0	0	0	0	20	20	100.00%
Total	20	16	20	23	26	23	20	25	15	12	20	20	240	80.83%

A=Direct Red C-380 B=Reactive Red 120 C=Reactive Red 123 D=Reactive Red 195	E=Reactive Red 2 F=Reactive Red 228 G=Direct Red 84 H=Reactive Red 180	I=Reactive Red 198 J=Reactive Red 239/241 K=Vat Red 10 L=Vat Red 15

4.2.3. First Derivatives

Another preprocessing technique was explored which involved taking the first derivative of each spectrum. This was performed to see whether the first derivative could help improve the discrimination between the dyed fibers. DA was performed assigning each Dye as its own class. This resulted in an 85.83% classification accuracy. The confusion matrix shown in Table 4.2.4 showed that Dyes A, C, G, and L were correctly classified 100.00% of the time. Dye K was classified correctly 99.00% of the time. Dyes B, E, and F had better discrimination, with Dyes B and E at 99.00% and Dye F and 98.00%. However, Dyes H, I, and J had poor classification accuracies. A t test was performed to determine if there was a significant difference between the means of the two data sets at the 0.05% confidence level. The result showed that there was no statistically significant difference (Table 4.2.3). The use of first derivatives as a preprocessing technique improved the discrimination of Dyes B, E, and F.

Table 4.2.3 t test results for different preprocessing techniques

	Normalized	First Derivative
Mean	0.850	0.858
Variance	0.0294	0.0850
Observations	12	12
Pearson Correlation	-0.1089	
Hypothesis mean difference	0	
df	11	
t stat	-0.0815	
P (T <= t) one tail	0.4682	
t critical, one tail	1.7959	
P (T <= t) two tail	0.9365	
t critical, two tail	2.201	

Table 4.2.4 DA confusion matrix for first derivatives

Predicted Dye / Actual Dye	A	B	C	D	E	F	G	H	I	J	K	L	Total	% Correct
A	100	0	0	0	0	0	0	0	0	0	0	0	100	100.00%
B	0	99	0	0	1	0	0	0	0	0	0	0	100	99.00%
C	0	0	100	0	0	0	0	0	0	0	0	0	100	100.00%
D	0	0	0	83	0	0	0	0	0	17	0	0	100	83.00%
E	0	0	0	0	99	0	0	0	0	1	0	0	100	99.00%
F	0	0	0	0	0	98	0	0	1	1	0	0	100	98.00%
G	0	0	0	0	0	0	100	0	0	0	0	0	100	100.00%
H	0	0	0	0	0	0	0	42	58	0	0	0	100	42.00%
I	0	0	0	0	0	10	0	80	10	0	0	0	100	10.00%
J	0	0	0	0	0	0	0	0	0	100	0	0	100	67.00%
K	0	0	0	1	0	0	0	0	0	0	99	0	100	99.00%
L	0	0	0	0	0	0	0	0	0	0	0	100	100	100.00%
Total	100	99	100	84	100	108	100	122	69	119	99	100	1200	85.83%

A=Direct Red C-380 B=Reactive Red 120 C=Reactive Red 123 D=Reactive Red 195	E=Reactive Red 2 F=Reactive Red 228 G=Direct Red 84 H=Reactive Red 180	I=Reactive Red 198 J=Reactive Red 239/241 K=Vat Red 10 L=Vat Red 15

4.2.4. Color Coordinates

Lastly, using the transmittance data, the color coordinates were calculated for each spectrum. The aim of this study was to see if utilizing color coordinates would increase the discrimination between the twelve dyed exemplars. This was analyzed using DA on all scans. For DA, all variables were used because the number of samples was greater than the number of variables. Ultimately, the confusion matrix showed that the classification accuracy was 65.25% (Table 4.2.6). However, the color coordinates placed the color of the fibers in the white region instead of the pink or red region. This was most likely the result of using the transmittance spectra instead of the reflectance, which is preferable. A t test was performed to determine if a significant difference occurred between the mean of the absorbance data and the mean of the color coordinate data. The results are shown in Table 4.2.5 and concluded that there was a statistically significant difference between the two means. Therefore, using color coordinates did not provide added discrimination between the dyes of the data set.

Table 4.2.5 t test for different analysis techniques

	Absorbance	Color Coordinate
Mean	0.850	0.6525
Variance	0.0294	0.0692
Observations	12	12
Pearson Correlation	0.7353	
Hypothesis mean difference	0	
df	11	
t stat	3.081	
P (T <= t) one tail	0.001451	
t critical, one tail	1.7959	
P (T <= t) two tail	0.002902	
t critical, two tail	2.201	

Table 4.2.6 DA confusion matrix for color coordinates

Predicted Dye / Actual Dye	A	B	C	D	E	F	G	H	I	J	K	L	Total	% Correct
A	87	0	0	0	0	10	0	0	0	0	3	0	100	87.00%
B	0	48	0	0	18	0	0	1	1	14	18	0	100	48.00%
C	0	0	89	0	0	0	0	0	0	0	2	9	100	89.00%
D	0	2	0	72	1	3	0	10	0	12	0	0	100	72.00%
E	1	31	0	2	23	9	0	4	9	12	9	0	100	23.00%
F	4	6	0	13	14	45	0	2	11	2	3	0	100	45.00%
G	0	0	0	0	0	0	95	0	0	0	0	5	100	95.00%
H	0	1	0	11	4	3	0	53	12	16	0	0	100	53.00%
I	0	1	0	11	18	21	0	9	31	9	0	0	100	31.00%
J	0	24	0	0	13	0	0	10	0	53	0	0	100	53.00%
K	0	7	0	0	4	0	0	0	0	0	87	2	100	87.00%
L	0	0	0	0	0	0	0	0	0	0	0	100	100	100.00%
Total	92	120	89	109	95	91	95	89	64	118	122	116	1200	65.25%

A=Direct Red C-380 B=Reactive Red 120 C=Reactive Red 123 D=Reactive Red 195	E=Reactive Red 2 F=Reactive Red 228 G=Direct Red 84 H=Reactive Red 180	I=Reactive Red 198 J=Reactive Red 239/241 K=Vat Red 10 L=Vat Red 15

4.3. All Dyed Exemplars Analyzed at ISP

The representative absorbance spectra for the twelve exemplars obtained using an instrument at the Indiana State Police are presented in Figure 4.3.1. Overall, these spectra resemble those obtained with the IUPUI instrument.

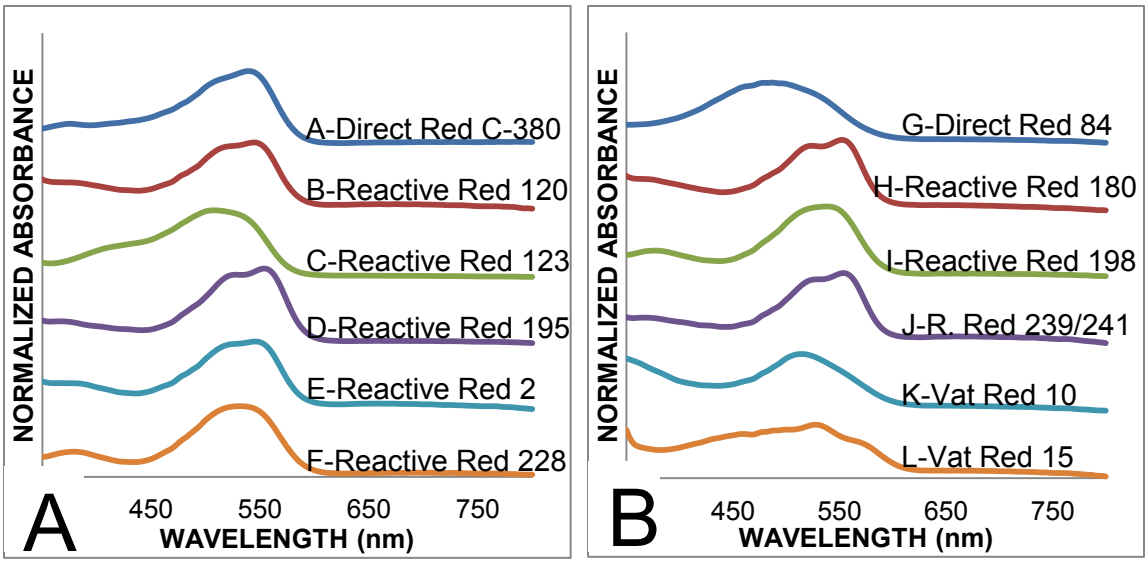


Figure 4.3.1 Representative spectra for all exemplars analyzed at ISP

The AHC dendrogram for fibers A-L is shown in Figure 4.3.2, with three distinct classes being formed based on the truncation line. The replicate fibers from Dyes C, G, K and L clustered well. These fibers have low levels of dissimilarity (C at 0.0033, G at 0.013, K at 0.013 and L at 0.033) proving that the replicates were similar. The replicate fibers from Dyes A, B, D, E, F, H, I, and J did not cluster as well and were split within the class. The top star indicates clustering between a fiber from Dyes D, E, and F with fibers from Dye J. The second star indicates clustering between fiber from Dyes B and E. The third star marks the clustering between fibers from Dyes H and J. The bottom star marks the clustering of fibers from Dyes B, E, F, and I. When comparing this to the IUPUI dendrogram, both data sets concluded that three classes were formed. However, there were small differences in the behavior of the replicates. The IUPUI data set was more behaved than the ISP data set. This is shown when

comparing the levels of dissimilarity. IUPUI had five dyes where all ten fibers clustered together and the level of dissimilarity ranged from 0.0023 to 0.0043. ISP only had four dyes where all ten fibers clustered together and the level of dissimilarity ranged from 0.0033 to 0.013. The central objects plot for this data set, Figure 4.3.3, showed that Class 1 contained two maxima wavelengths while the others contained a single maximum wavelength. Class 2 had a shifted maximum wavelength. Class 3 had a steady increase in absorbance until the maximum wavelength. This was comparable to the IUPUI data set.

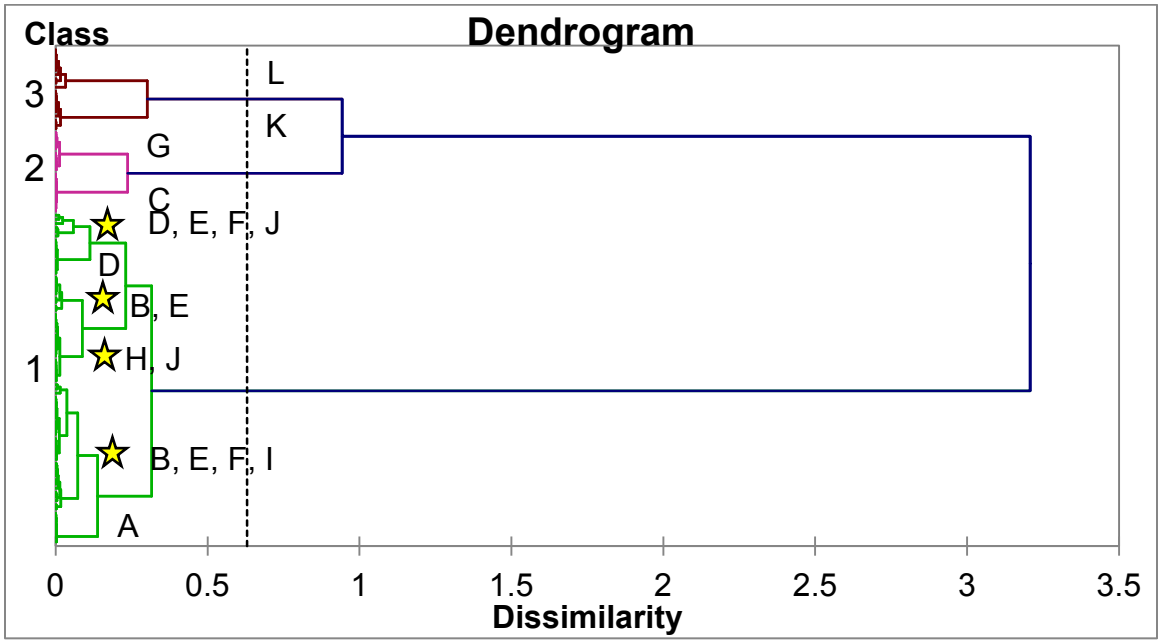


Figure 4.3.2 AHC dendrogram for all exemplars analyzed at ISP

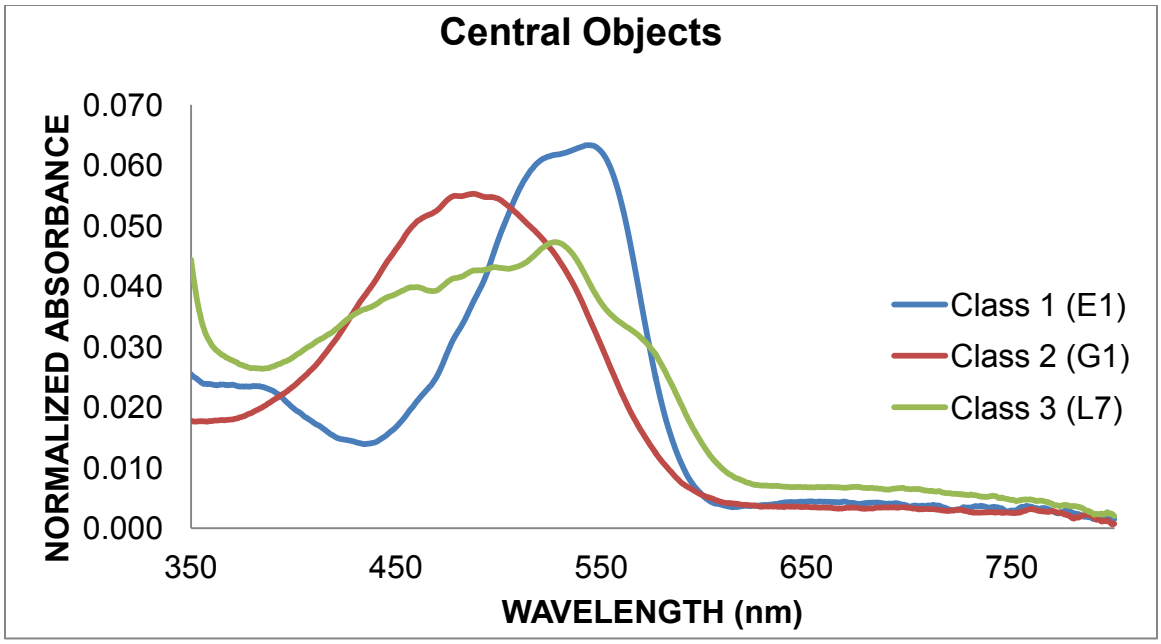
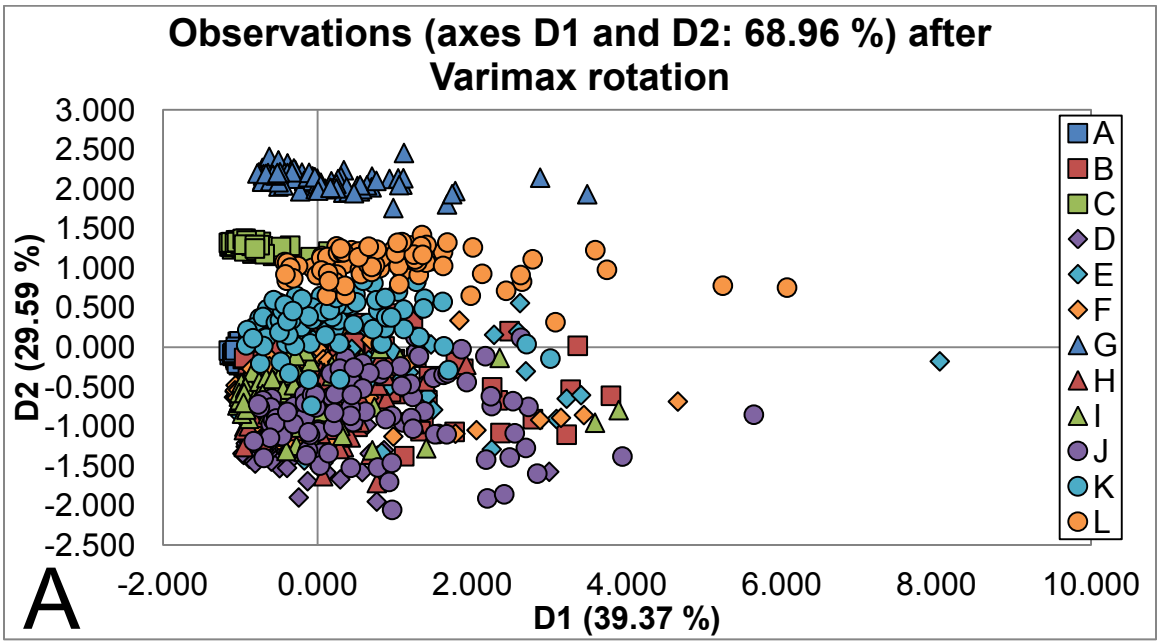


Figure 4.3.3 AHC central objects plot for all exemplars analyzed at ISP

The observations plot from PCA showed that the first two principal components captured 72.12% of the total variance (data not shown). Class 2, containing Dyes C and G, were readily distinguishable from the other dyes and each other. Class 3, containing Dyes K and L, were also distinguishable from the remaining dyes but not quite distinguishable from one another. The remaining dyes were not distinguishable. To obtain a better view of the dimension of the data, PCA was performed utilizing Varimax rotation by 3 factors. The total variance captured decreased to 68.96%. Now, Dyes C, G, K, and L were clearly distinguished from the other dyes. This can be seen in Figure 4.3.4A, which is a plot of the first two rotated principal components. The data is coded according to dye. Dyes B, E, F, H, I, and J are still overlaid indicating it is harder to distinguish between these dyes. Figure 4.3.4B is a plot of the second and third rotated principal components. Therefore, when the third dimension is considered Dye A was distinguishable from the others. The formation of two groups within the remaining dyes was seen. One group contained Dyes B, E, F, and I. The other group contained D, H, and J. When comparing the two data sets, more variance was captured in the first two principal components for the ISP data set.

Otherwise, both data sets showed that Dyes A, C, G, K, and L were distinguishable from the others dyes. Dyes B, E, F, and I were indistinguishable. Lastly, Dyes D, H, and J were indistinguishable.



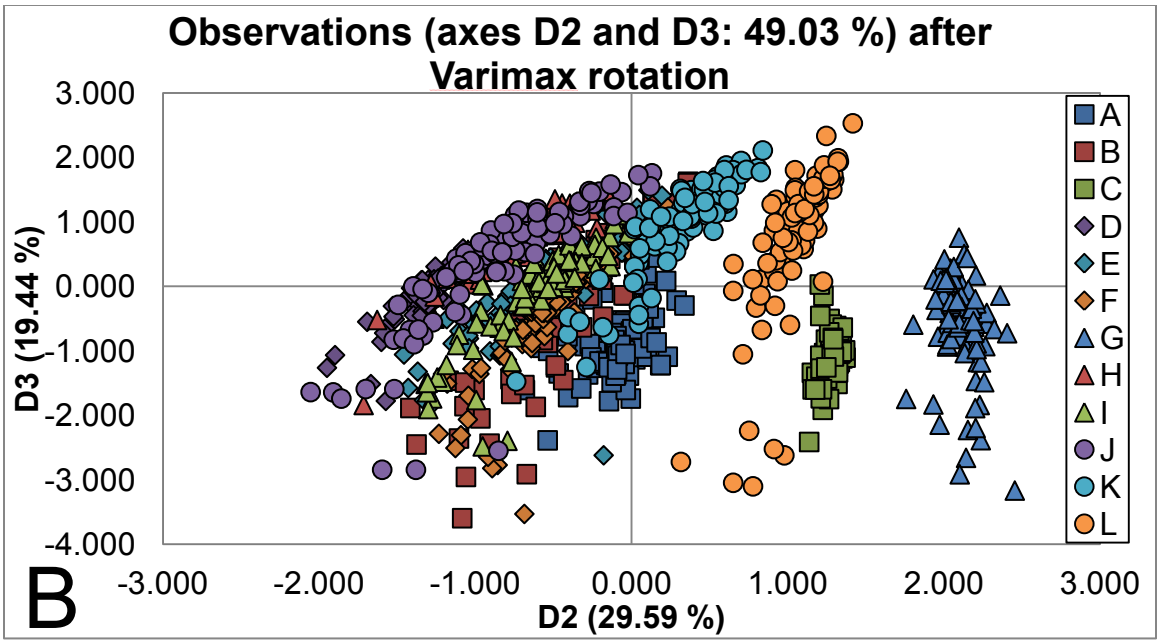


Figure 4.3.4 PCA observations plot for all exemplars analyzed at ISP

The correlation regions, determined from the factor loadings, overlaid with the representative spectra for this data set is shown in Figure 4.3.5. The main area of negative correlation for the first principal component occurred between 497-517 nm and the main area of positive correlation occurred between 605-770 nm. The negative correlation range included the maximum wavelength of Dyes C and K. The positive correlation range included the tails after the maximum wavelengths. The main area of positive correlation for the second principal component occurred between 412-494 nm; the main area of negative correlation occurred between 523-573 nm. The positive correlation range included the maximum wavelength of Dye G and the rising slope before the maximum wavelengths of the other dyes. The negative correlation range separated the single maximum wavelength dyes from the double maxima dyes. Overall, these ranges were consistent with the ranges determined by the instrument at IUPUI.

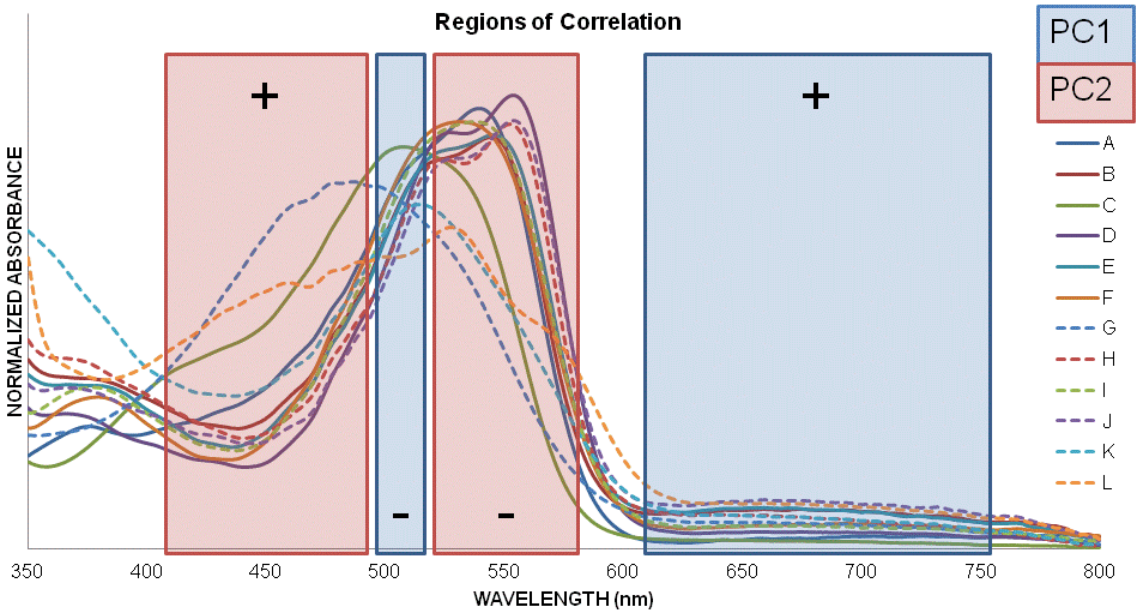


Figure 4.3.5 PCA regions of correlations for all exemplars analyzed at ISP

DA was performed in the same way for this data set as for the IUPUI set. When the fibers were assigned according to AHC class, the overall classification accuracy was 96.42% (data not shown). Class 2, which contained Dyes C and G were classified correctly 100% of the time. The second DA classified according to dye. The observations plot is shown in Figure 4.3.6. The classification accuracy was 81.25%. The most confusion occurred between fibers from Dyes B, E, F, and I. There was also confusion between fibers from Dyes D, H, and J. Dyes C, G, K, and L were correctly classified 100% of the time. The confusion matrix is shown in Table 4.3.1. Lastly, each fiber was assigned to a class totaling one hundred and twenty groups. The classification accuracy was 25.50% (data not shown). When simplified, it was seen that Dyes A, C, G, K and L were only misclassified as fibers from the same dye. Fibers from Dye B were misclassified as fibers from Dye E, F, and I 33.00% of the time. Dye E fibers were misclassified as fibers from Dyes B and I 56.00% of the time. Fibers from Dye F were misclassified as fibers from Dyes B, E, and I 23.00% of the time. Dye I fibers were misclassified as Dye E and F fibers 29.00% of the time. Fibers from Dye D were misclassified as fibers from Dye J 9.00% of the time. Dye H fibers were misclassified as fibers from Dyes D and J 4.00% of the time. Lastly, fibers from Dye J were misclassified as fibers from Dyes d and H 19.00% of the time. The new classification accuracy was 85.50%. When compared to the IUPUI data set, the same trend is seen where Dyes B, E, F, and I were most confused with one another while Dyes D, H, and J were most confused with one another. A t test was performed on the twelve class confusion matrix results for the two laboratories at the 0.05 confidence level. It was determined that there was not a statistically significant difference between the means of the two instruments. The t test results are found in Table 4.3.2.

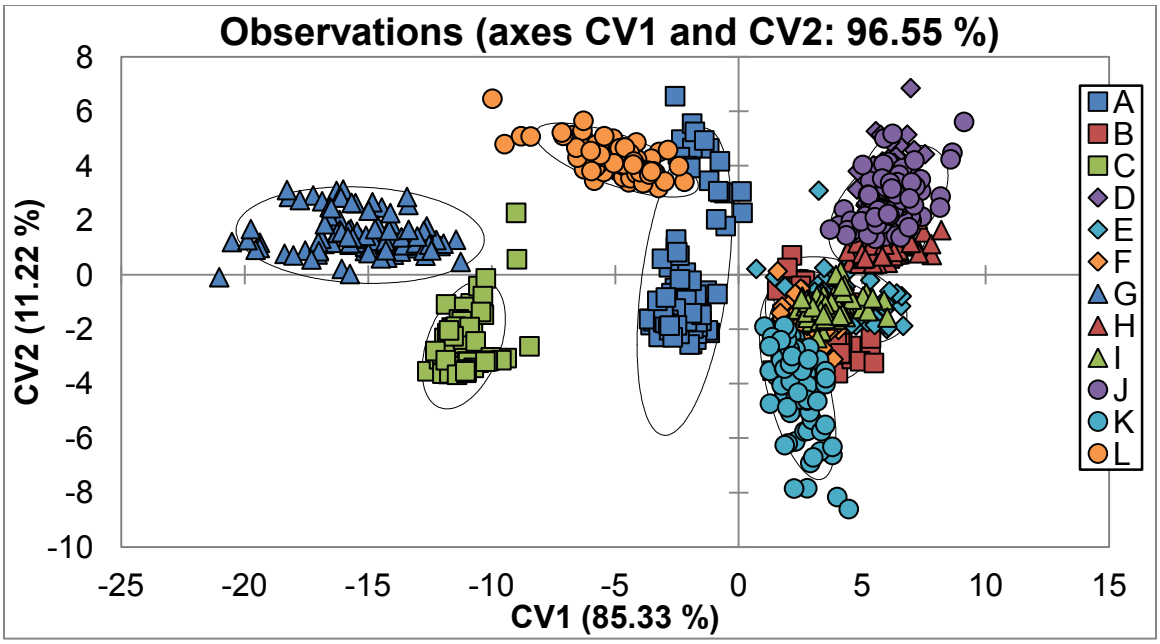


Figure 4.3.6 DA observations plot for all exemplars analyzed at ISP

Table 4.3.1 t test results for inter-laboratory study

	IUPUI	ISP
Mean	0.850	0.8125
Variance	0.0294	0.0446
Observations	12	12
Pearson Correlation	0.8629	
Hypothesis mean difference	0	
df	11	
t stat	1.211	
P (T <= t) one tail	0.1257	
t critical, one tail	1.796	
P (T <= t) two tail	0.2514	
t critical, two tail	2.201	

Table 4.3.2 DA confusion matrix for all exemplars analyzed at ISP

Predicted Dye / Actual Dye	A	B	C	D	E	F	G	H	I	J	K	L	Total	% Correct
A	92	0	0	0	0	0	0	0	0	0	0	8	100	92.00%
B	0	61	0	0	15	10	0	0	10	0	4	0	100	61.00%
C	0	0	99	0	0	0	0	0	0	0	0	1	100	99.00%
D	0	0	0	93	0	0	0	0	0	7	0	0	100	93.00%
E	0	22	0	0	39	0	0	3	35	1	0	0	100	39.00%
F	0	16	0	0	2	77	0	0	5	0	0	0	100	77.00%
G	0	0	1	0	0	0	99	0	0	0	0	0	100	99.00%
H	0	0	0	3	1	0	0	94	0	2	0	0	100	94.00%
I	0	7	0	0	15	8	0	0	70	0	0	0	100	70.00%
J	0	0	0	21	0	0	0	27	0	52	0	0	100	52.00%
K	0	1	0	0	0	0	0	0	0	0	99	0	100	99.00%
L	0	0	0	0	0	0	0	0	0	0	0	100	100	100.00%
Total	92	107	100	117	72	95	99	124	120	62	103	109	1200	81.25%

A=Direct Red C-380 B=Reactive Red 120 C=Reactive Red 123 D=Reactive Red 195	E=Reactive Red 2 F=Reactive Red 228 G=Direct Red 84 H=Reactive Red 180	I=Reactive Red 198 J=Reactive Red 239/241 K=Vat Red 10 L=Vat Red 15

Finally, the F-value plot revealed that the regions of most discrimination occur between 472-494 nm and 552-568 nm. When these regions are overlaid with the representative spectra, shown in Figure 4.3.6, the most discrimination occurred by the maximum wavelength of Dye G and on the decreasing slope of the other maximum wavelengths. These ranges are consistent with the ranges determined from the IUPUI data set. This showed that the most discriminating regions of the spectra were the increasing and decreasing slopes to either side of the maximum wavelengths.

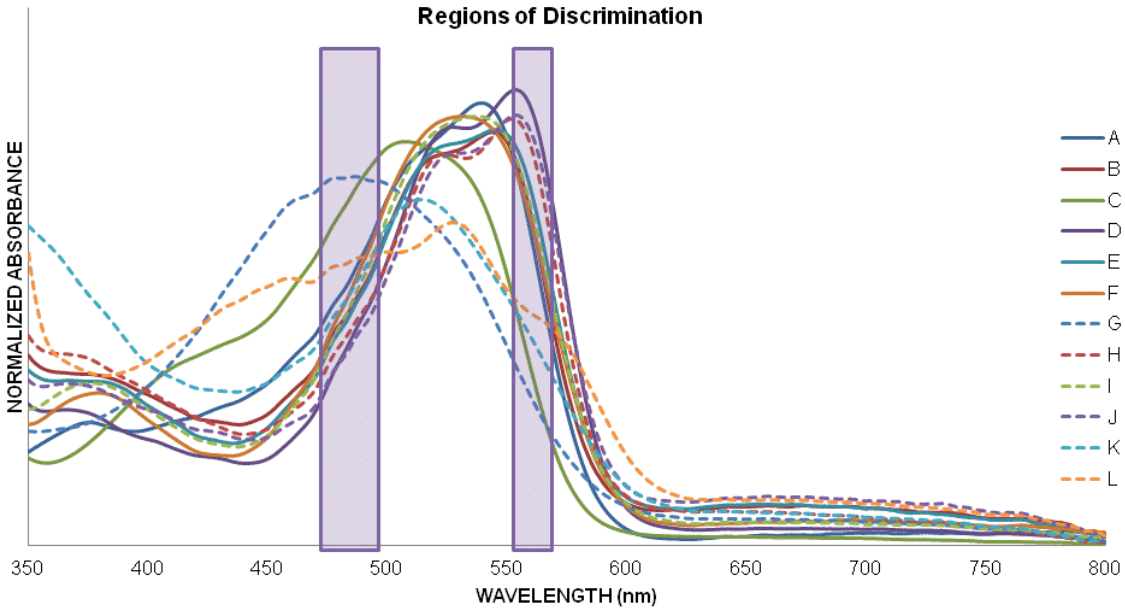


Figure 4.3.7 ANOVA regions of discrimination for all exemplars analyzed at ISP

CHAPTER 5. CONCLUSIONS

During the initial data collection, it was attempted to obtain the spectra of the different shades along the single fiber. This impacted the results of the chemometrics techniques. It was then noted that the SWGMAT guidelines suggest analyzing similar shades along a single fiber. Color bleaching occurred while the fibers were stored in glycerin for an extended period of time. Therefore, for long term storage fibers should be mounted in a more permanent medium (i.e. Permount). Lastly, using percent transmittance chould have affected the results of the color coordinate study.

The Testfabrics, Inc. samples were classified into four distinct groups with Reactive Red 123 and Reactive Red 195 being unique. This was also confirmed with PCA and DA analysis. Dyes B, E, and F were most often confused with one another and therefore, were not unique. It was noted that the regions of most discrimination were on the rise and fall but did not contain the maxima wavelengths. Lastly, a high degree of classification was seen within the external validation samples.

The dendrogram for all twelve exemplars analyzed at IUPUI showed that Direct Red C-380, Reactive Red 120, Direct Red 84, Vat Red 10, and Vat Red 15 were unique as supported by the results of PCA and DA. It was also concluded that Reactive Red 120, Reactive Red 2, Reactive Red 228, and Reactive Red 198 were more similar to each other than any of the other dyes as seen throughout the chemometric results. The same goes for Reactive Red 195, Reactive Red 180, and Reactive Red 239/241. The most discriminating regions again were on either side of the maxima wavelengths except for the maxima of

Direct Red 84. Lastly, utilizing the first derivative and color coordinates did not increase the discrimination between dyes.

The results for the exemplars analyzed at ISP were more or less identical to the results obtained at IUPUI. Small differences were seen between the data sets but overall the results were more similar. Utilizing a t-test between the results of the confusion matrix for the two data sets at 95% indicated that there was not a statistical significant difference. Therefore, it was shown that MSP is an acceptable technique for the analysis of dyed fibers that appear to be visually similar. It was also able to discriminate between red cotton fibers visually different. This will also be true when applied to other fiber types (natural or synthestic).

PART 2: TAPE BACKING ANALYSIS

CHAPTER 6. INTRODUCTION

The aim of this study was to evaluate X-ray fluorescence spectrometry and direct analysis in real time/mass spectrometry for the analysis of electrical tape backing. Should these nondestructive techniques be used in everyday forensic tape analysis? The XRF analysis was requested by contacts at the Alcohol, Tobacco, Firearms, and Explosives Laboratory. The DART/MS analysis was requested by contacts at the Federal Bureau of Investigation Forensic Laboratory. Each study was achieved in two parts. The first part was the discrimination of the black adhesive tape samples. The second part was the discrimination of colorless adhesive samples. However, DART/MS was performed in positive ion mode, negative ion mode, and contained samples with tinted adhesives. Therefore, the study consisted of six parts. The overall, evaluation included four chemometric techniques: AHC, PCA, DA, and ANOVA.

6.1. Tape Composition

Tapes often find their way into the forensic laboratory for identification and comparison. In the criminal world tapes are used as ligatures, restraints, packaging, and to construct improvised explosive devices. There are many classes of tape (e.g., duct tape, electrical tape, packaging tape, office tape, and masking tape).[56] Tapes also differ in design and chemical composition within a given class, which makes tape analysis feasible in the forensic laboratory. However, differences can also be found between different rolls made by the same manufacturer.[56]

The two most important features of tape are the chemical composition of the adhesive and the backing. The adhesive can be composed of a single component or a multicomponent system. In a single component system, it is common to see acrylic-based adhesives but silicone-based adhesives are also in use. In a multicomponent system, the adhesive consists of a blend of elastomers and resin tackifiers. Elastomers are natural or synthetic rubbers. The most common natural rubber is cis-polyisoprene. Synthetic rubbers are co-polymers like styrene-butadiene or styrene-isoprene-styrene and styrene-butadiene-styrene. Tackifiers lower the glass transition temperature and viscosity of the elastomer. Tackifiers can be aliphatic C5 hydrocarbons, polyterpenes, or natural rosin esters. Fillers, stabilizers, extenders, and fire retardants may also be added to the blend making a more complex system. The backing can either be made of fabric, polyester, cellulose acetate, polyethylene, polypropylene, or poly (vinyl chloride). Plasticizers are used to extend the flexibility of plastics like PVC. The most common plasticizers are phthalates, sebacates, adipates, aromatic oils, or aliphatic oils.[56]

The rest of the discussion will focus on electrical tape because the study was performed using electrical tapes. The backing consists of polyvinyl chloride. Dioctyl phthalate is the most common plasticizer; however, aliphatic sebacates and adipates may also be seen. The black color is due to the addition of carbon black. The backing can also contain fillers such as calcite, rutile, kaolinite, and talc. Antimony oxide, phosphates, or chlorinated hydrocarbons are added as fire retardants. Stabilizers may be added, which includes but is not limited to lead carbonate, lead sulfate, calcium, lead, cadmium, and barium stearates, dibutyl tin, and diphenyl urea. The adhesive can contain a variety of natural rubber and copolymer elastomers. The adhesive can be colorless or black (with the addition of carbon black). Also, plasticizers may be present in the adhesive that have “migrated” from the backing.[56]

6.2. Tape Analysis

The current practice for the forensic analysis of pressure sensitive tape standardized (PST) by the Scientific Working Group on Materials includes visual/microscopic examination and instrumental analysis.[57] Physical characteristics such as thickness, width, and adhesive color are determined with visual and microscopic examination. Instrumental analysis includes Fourier transform infrared spectroscopy (FTIR), scanning electron microscopy/energy dispersive spectrometry (SEM/EDS), and pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS). Other instrumental techniques can be utilized such as X-ray fluorescence (XRF), X-ray powder diffraction (XRD), and inductively coupled plasma (ICP). The organic composition of PST is studied by FTIR and Py-GC/MS. For FTIR, some of the organic constituents may be masked by the inorganic components; therefore, Py-GC/MS allows analysis of all organic comstituents. The inorganic components and elemental profiles are studied by SEM/EDS.[57]

6.2.1. X-Ray Fluorescence

One technique utilized in this study was XRF. The sample is irradiated with a primary beam of X-rays displacing an electron. This electron hole is filled by the movement of a higher orbital electron. The energy difference is released as X-ray photon. The X-ray beam can interact in other ways with the electrons of the samples. The primary beam hits the sample and some portion of the beam will go right through; this is the transmitted beam. The primary beam may also get scattered at an angle producing coherent and incoherent scattered wavelengths. Lastly, fluorescent radiation can occur.[58,59]

When the primary beam hits the sample continuous and characteristic radiation occurs. Continuous radiation results from the emission of a broad wavelength band of radiation and is caused by the deceleration of the primary electron due to the sample's electrons. Characteristic radiation occur when the primary electron contains greater energy than the sample's electron; ultimately,

ejecting the sample's electron as a photoelectron. The spectrum consists of a series of discrete wavelengths which are characteristic of the element. An example of this is shown in Figure 6.2.1.[60]

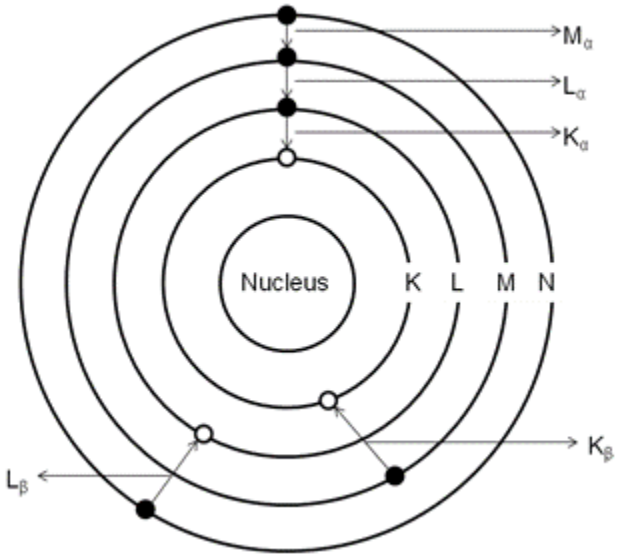


Figure 6.2.1 Transitions producing X-rays

An X-ray spectrometer contains five major components: excitation source, sample holder, disperser, detector, and data software. The most commonly seen excitation sources are the electron and the X-ray photon sources. A sealed X-ray tube containing an anode of Cr, Rh, W, Ag, Au, or Mo generates the high voltage that produces continuous and characteristic radiation. The detector converts the X-ray photon energy into voltage pulses. The most commonly seen energy dispersive detector is the Si(Li) detector. The X-ray photons interact and produce a number of electron hole pairs as given by Equation 6.1 where n is the number of electron hole pairs, E is the energy, and e_i which is the average energy required to produce one pair equals approximately 3.8 eV.

$$n = \frac{E}{e_i} \qquad \text{Equation 6.1}$$

The resolution, R, for this type of detector is calculated using Equation 6.2 where σ_{noise} is the noise contribution which is around 100 eV, F is the Fano factor which is around 0.12.[58,59]

$$R = \sqrt{[(\sigma_{noise})^2 + \left(2.35\sqrt{e_i F E}\right)^2]} \qquad \text{Equation 6.2}$$

In forensic science, XRF has been used to analyze pigments in paint samples and inks[61], atmospheric contaminants, counterfeit coins[62,14], toners[63], and gunshot residue.[64,65] Worley *et al*. imaged visible and latent fingerprints by detecting the inorganic constituents with micro-x-ray fluorescence methods.[66]

The same set of samples that were analyzed in this study was also used to study whether electrical tapes can be differentiated by surface texture, elemental composition, and organic composition of the backing.[15] It was concluded that the surface texture within a given roll is highly reproducible and that the surface texture between rolls of some brands can be differentiated by year of manufacture. The elemental composition was determined by SEM/EDS and evaluated with chemometric techniques such as AHC, PCA, and DA. One major finding was that most 3M tapes with black adhesive can be discriminated based on the absence or presence of lead, as well as relative amounts of other elements such as magnesium, aluminum, silicon, antimony, and calcium. Important exceptions were Super 88 and Super 33+, whose elemental profiles proved to be so similar as to be nearly indistinguishable. Another major finding is that clear tape adhesives can be differentiated into two categories: those with aluminosilicate filler and those without. Lastly, changes in the formulation within a brand can serve as a discriminating factor for year of manufacture. Overall, the 67 rolls of tape could be separated into 36 differentiable groups.[15] The organic composition was examined using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy (ATR-FTIR), evaluated with chemometrics, and compared to results obtained with SEM/EDS.[16] Overall, it was concluded that FTIR analysis provided no additional advantage to discrimination of the tape samples than SEM/EDS provided, meaning that brands that could not be differentiated by SEM/EDS could also not be differentiated by FTIR. However, FTIR provided better discrimination between the adhesives of the tape samples than SEM/EDS provided.[16] The inorganic components of tape samples make it suitable for XRF analysis and the data obtained is suitable for chemometric analysis.

6.2.2. Direct Analysis in Real Time/Mass Spectrometry

The other technique utilized in this study was DART/MS. DART is a new ionizing technique for mass spectrometry that is growing in use in the forensic laboratory. DART was designed to overcome the limitations of the traditional ionizing techniques such as requiring the sample to be introduced into a system under vacuum and a shielded source to protect the operator. Other ion sources expose the sample to elevated temperatures, laser radiation, ultraviolet irradiation, solvents, or electrical discharges. When coupled to a time-of-flight mass spectrometer, DART provides rapid analysis of solid, liquid, or gaseous materials at ambient pressure. Accelerants, drugs of abuse, dyes, explosives, fibers, inks, and spices are just a few of the many compounds that have been analyzed with DART/MS.[67,68,69]

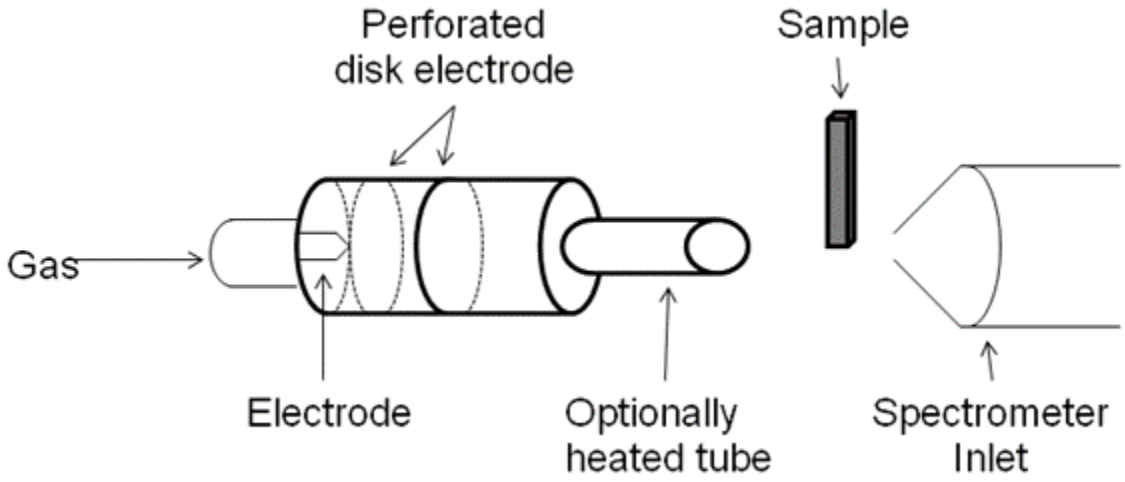


Figure 6.2.2 DART/MS instrumentation

The instrument setup is portrayed in Figure 6.2.2. The mechanisms behind DART ionization depend on the interactions of excited-state species with the sample and atmospheric gases. The excited-state species are generated when the gas (helium, nitrogen, or neon) flows through a chamber with an electrical charge creating glow discharge, which contains ions, electrons, and excited-state species. The charged species are removed when the glow discharge flows through a grid or perforated lens leaving neutral molecules, including excited-state atoms/molecules. The excited-state species will then

react with atmospheric gases and the sample.[67,68,69] If nitrogen or neon is used as the gas, Penning ionization is the primary reaction. This reaction is shown in Equation 6.3 where M* is the metastable atom and S is the analyte.

$$M^* + S \rightarrow S^{+\cdot} + M + e^- \qquad \text{Equation 6.3}$$

The metastable atom transfers energy to the analyte releasing an electron and producing a radical cation. When helium is used as the gas then positive-ion formation is the primary reaction. The mechanism for this type of reaction is shown in Equation 6.4.

$$He(2^3S) + H_2O \rightarrow H_2O^{+\cdot} + He(1^1S) + e^- \qquad \text{Equation 6.4A}$$

$$H_2O^{+\cdot} + H_2O \rightarrow H_3O^+ + OH^{\cdot} \qquad \text{Equation 6.4B}$$

$$H_3O^+ + nH_2O \rightarrow [(H_2O)nH]^+ \qquad \text{Equation 6.4C}$$

$$[(H_2O)nH]^+ + S \rightarrow SH^+ + nH_2O \qquad \text{Equation 6.4D}$$

Equation 6.4A involves the transfer of energy from the excited state helium atom to atmospheric water producing an electron and a water radical cation. The water radical cation reacts with atmospheric water to produce the hydronium ion and a hydroxyl radical as shown in Equation 6.4B. The hydronium ion reacts with more atmospheric water to form a ionized water cluster (Equation 6.4C) which then transfers a proton to the analyte producing an cation (Equation 6.4D). If desired, negative-ion formation is possible for all gases and involves surface Penning ionization.

$$M^* + surface \rightarrow M + surface + e^- \qquad \text{Equation 6.5A}$$

$$e^-_{fast} + gas \rightarrow e^-_{slow} \qquad \text{Equation 6.5B}$$

$$e^-_{slow} + O_2 \rightarrow O_2^- \qquad \text{Equation 6.5C}$$

The first step in the reaction (Equation 6.5A) involves the reaction of the metastable atom with the surface of the sample to produce an electron. Equation 6.3B involves the thermalization of the electron by atmospheric gases. Lastly, the thermalized electron undergoes electron capture with atmospheric oxygen producing an superoxide (Equation 6.3C).[68] Therefore, the resultant mass spectrum will either show a protonated analyte or analyte anion depending of the type of reaction occurring and mode of detection.[67,68,69]

Mehltretter *et al.* have conducted a two part study on the discrimination of ninety black electrical tapes (the same samples used in this study) by stereomicroscopy, FTIR, SEM/EDS, and pyrolysis gas chromatography/ Py-GC/MS. The first part of the study was to examine the adhesives. They found that there were three different colors of adhesives by stereomicroscopy; clear/colorless, clear with brown tint, and black. This difference in adhesive color was also corrected with FTIR analysis which also separated the groups further by chemical composition into eight groups (six for clear adhesives and two for black adhesives) resulting in two unique samples. Py-GC/MS confirmed the eight groups formed by FTIR and separated them further into sixteen groups (twelve for clear and four for black) resulting in seven unique samples. SEM/EDS analysis only separated the rolls into five groups resulting in two unique samples. Therefore, Py-GC/MS showed the best discrimination of the adhesives. When all the individual techniques were considered together, it was concluded that the ninety samples can be discriminated into eighteen groups resulting in eight unique samples.[70]

The second part of the study examined the backings of the same ninety rolls of electrical tape. Stereomicroscopy showed twenty unique samples based on width and backing thickness. The remaining seventy samples were separated into four groups. FTIR analysis resulted in seven unique samples with the remaining eighty-three samples separating into seven groups to give a total of fourteen groups. Py-GC/MS analysis resulted in five unique samples with the remaining samples separating into seven groups for a total of twelve groups. SEM/EDS separated the samples into fifteen groups with three unique samples. SEM/EDS showed to be the best at discriminating the tape backings. When all techniques were considered, it was concluded that the ninety samples can be discriminated into forty groups with thirty-one unique samples. Finally, when the adhesive and backing data was considered together, forty-four groups were distinguished with thirty-three unique samples. Overall, better discrimination was seen when all the data was considered.[71]

CHAPTER 7. MATERIALS AND METHODS

7.1. XRF Materials

The ATF electrical tape reference collection currently includes over seventy rolls of tape representing over thirty nominal brands. The rolls analyzed in this study are listed in Table 7.1.1 along with the manufacturer (if known), origin, and Underwriters Laboratory (UL) certification number. Overall, thirty-three brands representing six known manufacturers were analyzed. All of the same brands were analyzed as published by Goodpaster *et al*.[15,16] However, not all of the rolls previously analyzed were analyzed here.

Table 7.1.1 Electrical tape samples analyzed by XRF

Brand	Roll(s)	Origin	UL	Manufacturer
Scotch Super 33+ Cold Weather	3M1	U.S.A.	539H	3M
Scotch 33	3M2A, 3M2B, 3M2C	U.S.A.	539H	3M
Scotch Super 88	3M3A, 3M3B, 3M3C	U.S.A.	539H	3M
Scotch Super 33+ (2002 + ↑)	3M4A, 3M4B, 3MC	U.S.A.	539H	3M
Scotch Super 33+ (2001 + ↓)	3M4D, 3M4E	U.S.A.	539H	3M
Temflex 1700	3M6A, 3M6B, 3M6C	U.S.A.	539H	3M
Radio Shack (Temflex 1700)	3M7	U.S.A.	539H	3M
Tartan 1710	3M8A, 3M8B, 3M8C	Taiwan	9Z53	3M
ACE Hardware	ET1	Taiwan	362K E49341 #33180	ACHEM Technology
All-Temp	ET2	Taiwan	362K E49341 FR 101	ACHEM Technology
Auto Solutions	ET3	China	74HK E206648	ACHEM Technology
Bengal	ET4A, ET4B, ET4C, ET4D, ET4E			
Champion Quality Tools	ET5A, ET5B, ET5C, ET5D	China	57RJ E220094 #1920	Ningbo Sincere Adhesive Products
Duck	ET6A, ET6B, ET6C	Taiwan China	362K 74HK E49341	ACHEM Technology
Duck (668 Pro)	ET7	Taiwan	362K E49341	ACHEM Technology
Electro Tuff	ET8	Taiwan	362K E52811	ACHEM Technology
Frost King (ET60)	ET9A, ET9B	Taiwan	206T	Globe Industries
Frost King (ET60FR)	ET9C, ET9D	China	57RJ	Ningbo Sincere Adhesive Products

Table 7.1.1 Continued.

Brand	Roll(s)	Origin	UL	Manufacturer
Globe	ET10	Taiwan	206T E62265 #220	Globe Industries
GE	ET11	Taiwan	206T E62265 #230	Globe Industries
Intertape	ET12	Taiwan	362K E52811 #101	ACHEM Technology
LePage's	ET13	-	906B	-
Leviton	ET14	U.S.A.	-	-
Manco	ET15	Taiwan	590J	-
Michigan Industrial Tools	ET16	Taiwan	590J E50292 #33546	Symbio
Nashua A-7	ET17	U.S.A.	116A	-
Power First	ET18	China	98LJ E174965	Hebei Huaxia Enterprise
Permacel (AW 8 ½)	ET19	-	705B	-
Permacel (P-29 Plus)	ET20	-	705B	-
Radio Shack	ET21	Taiwan	362K E52811 #101	ACHEM Technology
Shurtape	ET22	-	859X	-
Powerworks	ET23	Taiwan	590J E50292 #33546	Symbio
Vanguard	ET24A, ET24B	U.S.A.	521D	-
WUL	ET25	Taiwan	590J #33546	Symbio

7.2. XRF Instrumental Analysis

Each sample was mounted on a carbon planchette prior to analysis. The instrument utilized was an Edax Eagle III μProbe X-ray fluorescence spectrometer (EDAX, Mahwah, NJ) with a solid-state silicon (lithium drifted) detector. The detector had high vacuum cryogenic assemble with a fixed beryllium window. The samples were irradiated with a rhodium anode X-ray tube with beryllium window and 22^{o} take-off angle. The X-ray capillary diameter was set at 300 μm and an X-ray incidence angle of 65^{o}. The data was collected at 30 kV for about 100 live seconds.

Peaks were generated for: aluminum, silicon, chlorine, calcium, antimony, titanium, iron, zinc, and lead.

7.3. XRF Data Analysis

After the data was collected it was separated according to the color of the adhesive (clear or black). Each set was normalized by dividing each peak area by the square root of the sums of squares of all peaks for a given sample. This corrects for variations in sample amount and instrument response. This was followed by autoscaling, which is mean subtraction (subtracting the mean for each variable) and variance scaling (dividing by the standard deviation for each variable). Autoscaling accounts for differences in precision between variables, so that equal weight is placed on each variable. Multivariate statistical techniques were then applied, including AHC, PCA, DA, and ANOVA, using XLSTAT2010 (Addinsoft SARL, Paris, France) a Microsoft Excel add-in. For all techniques, all replicate scans were considered individually. Unlike before, DA was performed on all variables because the number of samples was greater than the number of variables.

7.4. DART/MS Materials

The tape collection analyzed in this study consisted of ninety black electrical tape rolls from a number of brands. Table 7.4.1 is a summary of the products analyzed in this study. Some rolls of tape did not indicate the manufacturer on the label and it is common for tapes to be purchased from the manufacturer and sold under private labels. Therefore, the sample set was not selected to be representative of all manufacturers or used to provide information on an unknown tape sample's origin or source. All of the same brands that were previously studied by Mehltretter *et al*.[70,71] were studied here except for sample 75 (3M Scotch Electrical Rubber Splicing Tape) because of its greater thickness.

Table 7.4.1 Electrical tape samples analyzed by DART/MS

Roll Number	Brand Name	Product
1	Marcy Enterprises, Inc.	MA 750
2	Advance®	AT7, BS3924, 31/90Tp
3	Work Saver™	Stock No. 55, 5-Color P.V.C. Tape Assortment
4	Tesa	40201, No. 111 E52811A
5	Tape It, Inc.	E-60
6	Qualpack®	1346, 6-Color
7	Marcy Enterprises, Inc.	MA 750
8	Manco®	200 MPH, AE-66
9	Archer® (Packaged for Radio Shack)	64-2349
10	3M Scotch™	Super 88, 054007-06143
11	3M Scotch™	Super 33+, 10414 NA
12	3M Scotch™	Super 33+, 10455 NA
13	3M Scotch™	Super 33+
14	Frost King®, Thermwell Products Co., Inc.	ET60
15	3M Scotch™	Super 33+, 10455 NA
16	3M	Tartan 1710, part no. 054007-49656
17	3M Scotch™	Super 88, 054007-06143
18	3M Scotch™	Super 33+, Cat. 195NA
19	3M Scotch™	Super 33+, Cat. 194NA
20	3M Scotch™	Super 33+, 10414 NA
21	Manco®	P-66
22	Manco®	667 Pro Series
23	3M Scotch™	Super 88, 054007-06143
24	3M Scotch™	Super 88, 054007-06143
25	3M Scotch™	Super 33+, 054007-06132
26	3M Scotch™	Super 33+, 054007-06132
27	3M	Tartan 1710, part no. 054007-49656
28	3M	Tartan 1710, part no. 054007-49656
29	3M	Temflex 1700, 54007-69764
30	3M	Temflex 1700, 54007-69764
31	Regal®	Model ET-6
32	GE	GE2472-3DD
33	3M Scotch™	Cat. 190
34	3M	Tartan 1710, part no. 054007-49656
35	Frost King®, Thermwell Products Co., Inc.	ET60
36	3M	Tartan 1710, part no. 054007-49656
37	National	All-Purpose Grade
38	Manco®	P-660
39	3M Scotch™	Super 33+, 3744NA
40	3M	Tartan 1710, part no. 054007-49656
41	3M Scotch™	Super 33+, 200NA
42	National	All-Purpose
43	3M	Tartan 1710, part no. 054007-49656
44	3M	Tartan 1710, part no. 054007-49656
45	Calterm®	49605
46	Manco®	P-20
47	3M	Tartan 1710, part no. 054007-49656
48	Tape It, Inc.	36-T
49	Tape It, Inc.	36-T
50	General Electric	GE2472-31D
51	National	No. 101 E52811A
52	Frost King®, Thermwell Products Co., Inc.	ET60FR
53	National	No. 101 E52811A
54	3M Scotch™	Super 33+ (on core), 03404NA (on package)
55	Manco®	1219-60
56	Victor Products, Inc., Thermoflex	33-UL60, No. 101 E52811A
57	United Tape Company	UT-602
58	Frost King®, Thermwell Products Co., Inc.	ET60
59	Tuff Hand Tools	
60	Tuff Hand Tools	
61	3M Scotch™	88T
62	Nitto Denko	No. 228
63	3M Scotch™	Super 88, 054007-06143
64	3M Scotch™	Super 33+, 10455 NA
65	3M Scotch™	700 Commercial Grade, 054007-04218
66	L. G. Sourcing, Inc.	19453
67	Manco	P-66
68	3M Scotch™	Super 33+

Table 7.4.1 Continued.

Roll Number	Brand Name	Product
69	3M	Tartan 1710, part no. 054007-49656
70	Tyco Adhesives National Tape Products	No. 101, E52811A
71	Qualpack®	1346, 6-Color
72	Nitto Denko	No. 228
73	Frost King®, Thermwell Products Co., Inc.	ET60FR
74	3M Scotch®	700 Commercial Grade, 054007-04218
76	3M Scotch®	Super 33+, Cold Weather, 16736NA
77	3M Scotch®	Super 33+, 054007-06132
78	3M	Tartan 1710, part no. 054007-49656
79	3M Scotch®	700 Commercial Grade, 054007-04218
80	3M Scotch®	Super 88, 054007-06143
81	Ace (imported for Henkel Capital)	All Weather
82	Ace (imported for Henkel Capital)	Weather Resistant
83	3M Scotch®	Super 33+, 10414 NA
84	3M	Tartan 1710, General Use, 054007-49656
85	Frost King®, Thermwell Products Co., Inc.	ET60FR
86	Duck, Henkel Consumer Adhesives	Vinyl Electrical Tape
87	Nitto Denko	No. 21E
88	Frost King®, Thermwell Products Co., Inc.	ET60FR
89	Power Pro Craft	ETF
90	Duck, Henkel Consumer Adhesives	Extra wide electrical tape

7.5. DART/MS Instrumental Analysis

The adhesive was separated from the backing using hexane with a cotton swab/wipe and allowed to air dry. All samples were manually placed in-front of the JEOL IonSense DART (Saugus, MA) ionization stream which was coupled to a JEOL AccuTOF Mass Spectrometer, Model JMS-T100LC (Peabody, MA). Each roll was sampled twice, once each on two separate days, in triplicate. The data was collected and processed by JEOL DART Controller and Mass Center Main Software. The DART ionization source was operated in both the positive and negative ion mode with helium gas (99.99% purity) and the heating coil set at 400 ^{o}C. The electrode potentials were set at: orifice 1 at ± 30 V, orifice 2 at ± 5 V, ring lens at ± 5 V, and the glow discharge needle at ± 4000 V. Spectra were acquired for m/z 80 to 1000. For exact mass measurements, an external polyethylene glycol (Sigma) reference standard mass spectrum was taken with each data file. Standards of common plasticizers were also analyzed in both positive and negative ion mode. These standards were dioctyl phthalate (DOP), dioctyl adipate (DOA), di-n-nonyl phthalate (DNP), and trioctyl trimellitate (TOTM). Lastly, a mixture (SPEX CertiPrep) containing 2000 µL of DOP, DOA, butyl benzylphthalate (BBP), diethyl phthalate (DEP), dimethyl phthalate (DMP), and di-n-butyl phthalate (DNBP) was analyzed.

7.6. DART/MS Data Analysis

Once the data was obtained it was separated according to the color of the adhesive (clear colorless, clear with brown-tint, or black). The masses were rounded to the nearest whole number for consistency. For each of the three groups, a list of the five highest intensity masses was compiled. Preprocessing included normalization. It was found that autoscaling was unnecessary. Multivariate statistical techniques were applied with XLSTAT2010 (Paris, France), an Excel add-in. The techniques include AHC, PCA, DA, and ANOVA. The averages of the three scans were utilized for AHC so that the dendrogram would be readable. For PCA and ANOVA, all replicate scans were considered individual. The number of variables utilized in DA was determined when the cumulative percent of variance for the principal components reached 95%. This resulted in the use of 18, 4, and 12 principal components for clear-colorless, clear-tinted, and black adhesives, respectively.

CHAPTER 8. RESULTS AND DISCUSSIONS

8.1. XRF Black Adhesives

There were twenty rolls of black adhesive tape analyzed representing seven brands. These samples are listed in Table 8.1.1 along with their identifier.

Table 8.1.1 Black adhesive samples analyzed by XRF

Brand	Identifier
3M Super 33+ (Cold Weather)	3M1
33 Electrical Tape	3M2A-D
Super88	3M3A-C
Super 33+ (2001↑)	3M4A-C
Super 33+ (2001↓)	3M4D-E
700 Commercial Grade	3M5A-D
Temflex	3M6A-C
Radioshack Temflex	3M7

AHC placed these twenty rolls into three classes. The dendrogram (Figure 8.1.1) showed that 33 Electrical Tape and 700 Commercial Grade were placed into the same group; however, no confusion occurred between the samples. The samples of Super 33+ produced before 2001 were placed into their own group. Lastly, Super 33+ Cold Weather (CW), Super 88, Super 33+ produced after 2001, Temflex, and Radio Shack Temflex were placed in the same group. There was some confusion between Super 33+ CW and Super 88 indicated by the top star on the dendrogram. The bottom star indicates that confusion was also seen between Temflex and Radio Shack Temflex. When the dendrogram was examined in more detail, it is noticed that the four rolls of 700 Commercial Grade is split with relatively high dissimilarity (~50). The central objects plot in Figure 8.1.2 indicated that Class 1 contained calcium, antimony, and zinc. Class 2 contained antimony and a larger amount of lead. Class 3 contained calcium, antimony, zinc, and a larger amount of lead.

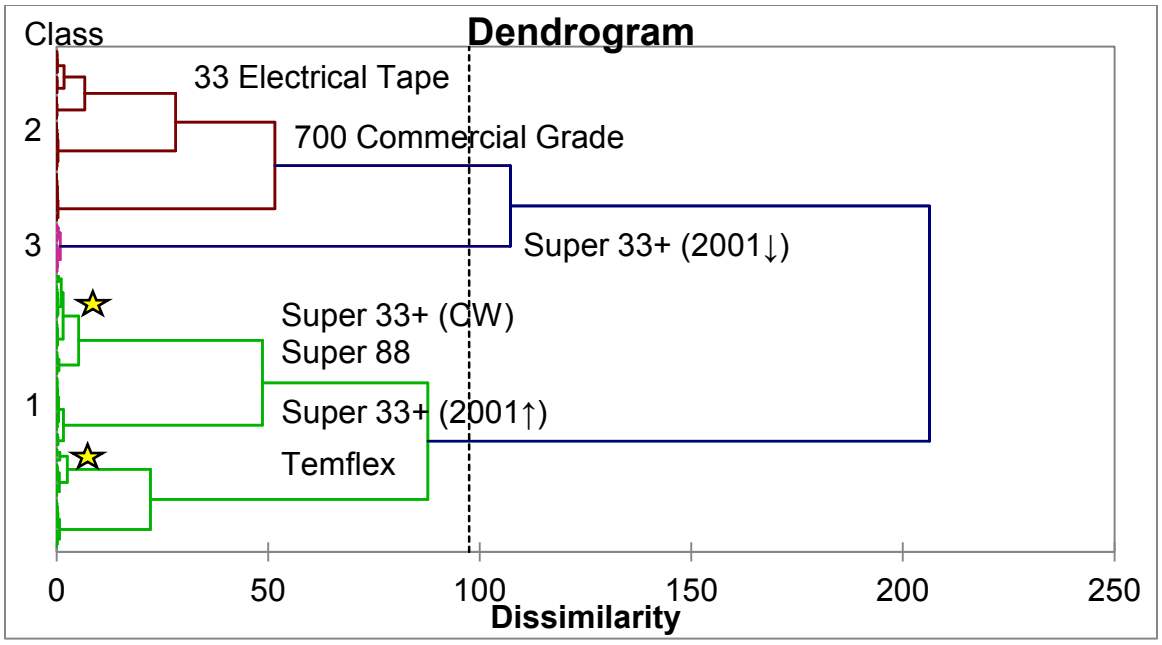


Figure 8.1.1 AHC dendrogram for black adhesives analyzed by XRF

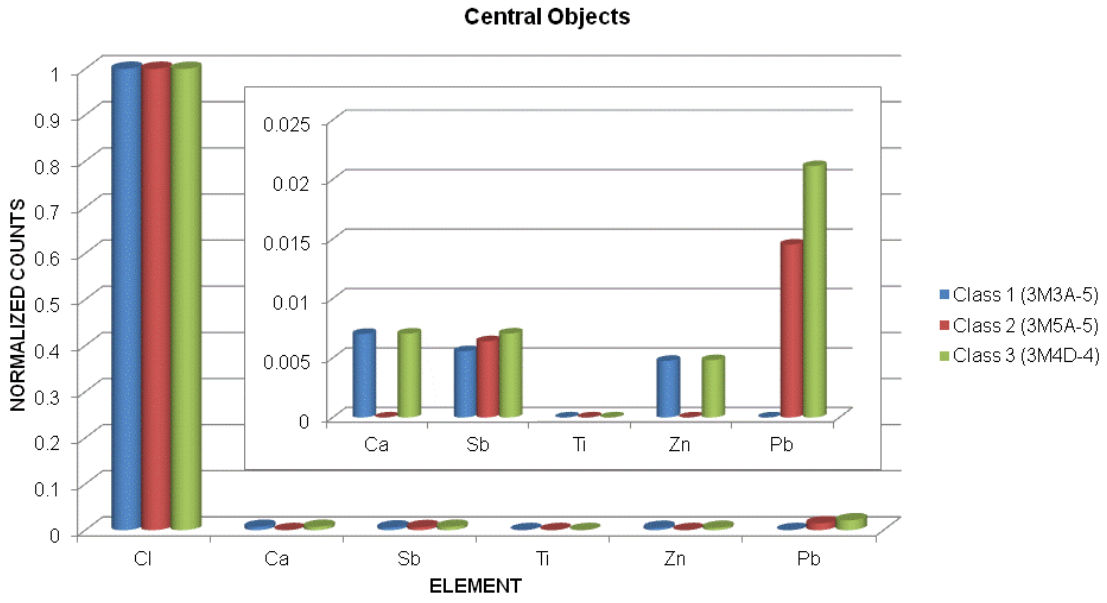
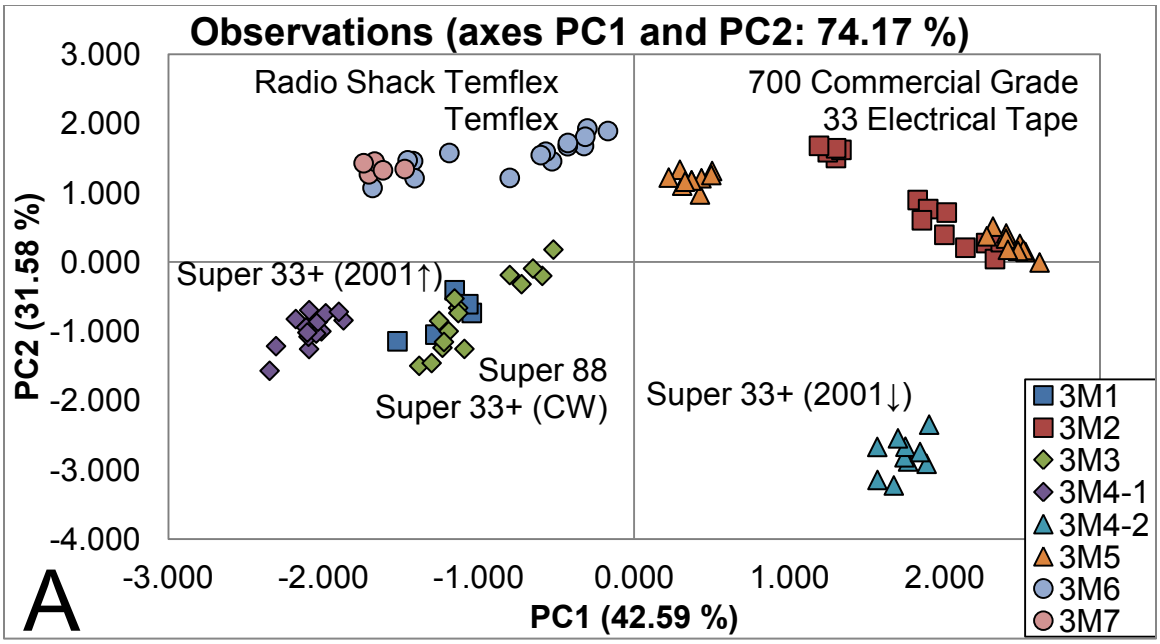


Figure 8.1.2 AHC central objects plot for black adhesives analyzed by XRF

PCA showed that in two dimensions 74.17% of the total variance was captured. When the observations plot was coded according to the three AHC classes it was noted that all classes were clearly distinct (data not shown). The observations plot coded according to brand showed that when the first two PCs are plotted against one another (Figure 8.1.3A), Temflex and Radio Shack Temflex were indistinguishable. This was also the case for Super 33+ CW and Super 88, as well as 33 Electrical Tape and 700 Commercial Grade. However, when the first and third PCs were plotted against one another (Figure 8.1.3B), 33 Electrical Tape and 700 Commercial Grade were distinguishable. Super 33+ CW and Super 88 were not distinguishable from one another. Temflex and Radio Shack Temflex were also indistinguishable.



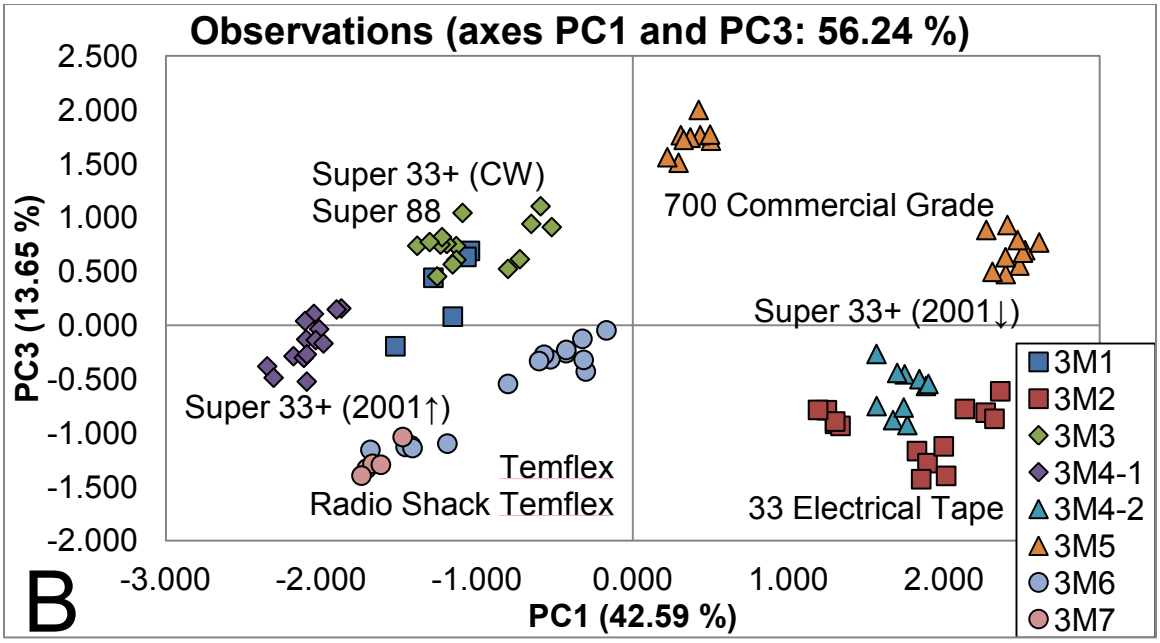


Figure 8.1.3 PCA observations plot for black adhesives analyzed by XRF

The factor loadings plot, Figure 8.1.4, indicated that the PC1 was positively correlated with lead and negatively correlated with zinc. PC2 was positively correlated with chlorine and negatively correlated with calcium. Lastly, the PC3 showed a positive correlation with antimony and no correlation with the other elements. Therefore, what separated 33 Electrical Tape and 700 Commercial Grade in Figure 8.1.3B was their amount of antimony.

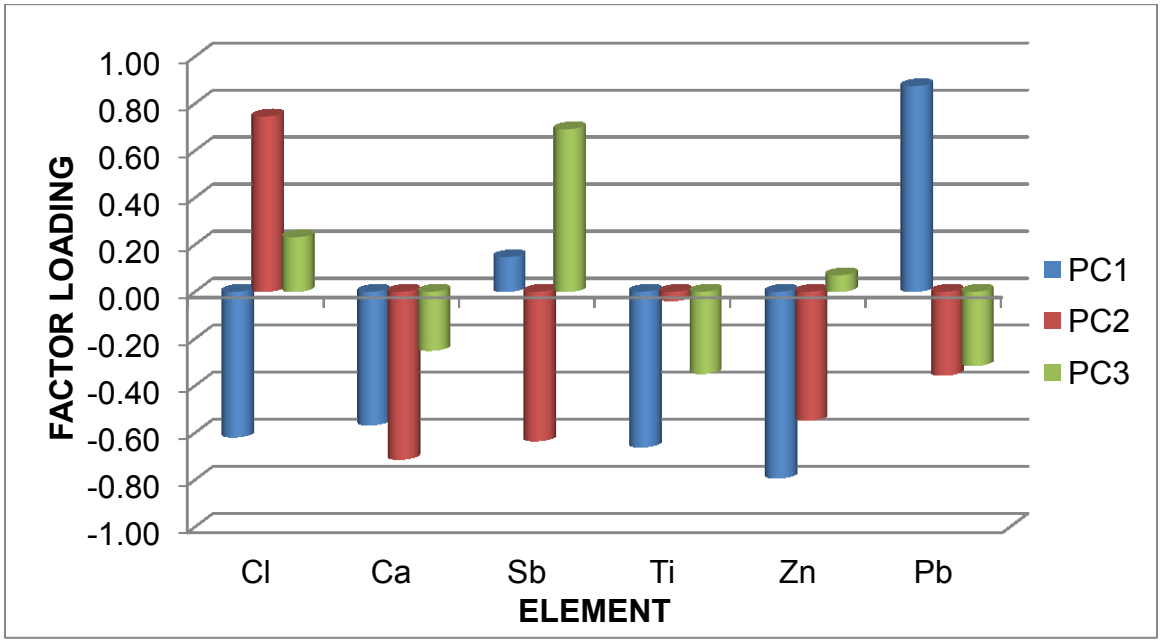


Figure 8.1.4 PCA factor loadings plot for black adhesives analyzed by XRF

Initial DA was calculated by placing the samples into one of three groups based on the clustering from AHC. The overall classification accuracy was 100.00% (data not shown). The next DA was calculated by placing the samples into groups according to brand and year, if necessary. The observations plot (Figure 8.1.5) showed that Temflex and 33 Electrical Tape overlapped. It also showed that Super 33+ CW, Super 88, and Super 33+ overlapped. However, the confusion matrix (Table 8.1.2) indicated that the overall classification accuracy was 96.94% where the only confusion occurred between Super 33+ CW and Super 88. The overlapping in the observations plot may have been caused by the low total variance captured in two dimensions being 83.68%.

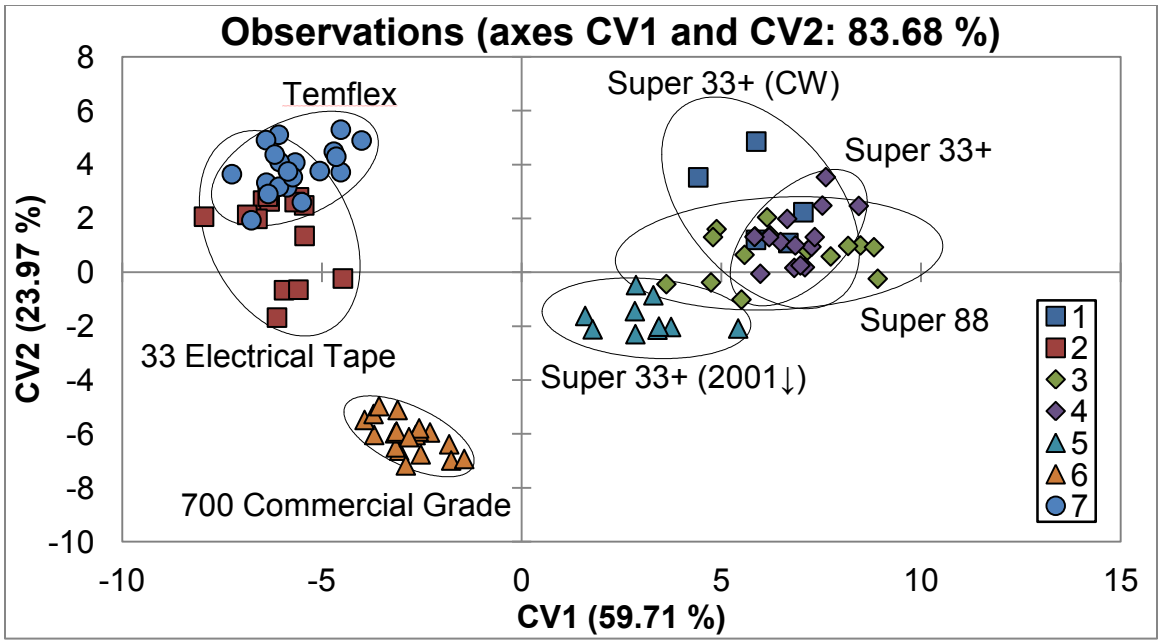


Figure 8.1.5 DA observations plot for black adhesives analyzed by XRF

Table 8.1.2 DA confusion matrix for black adhesives analyzed by XRF

Predicted Brand / Actual Brand	1	2	3	4	5	6	7	Total	% Correct
1	2	0	3	0	0	0	0	5	40.00%
2	0	14	0	0	0	0	0	14	100.00%
3	0	0	15	0	0	0	0	15	100.00%
4	0	0	0	14	0	0	0	14	100.00%
5	0	0	0	0	10	0	0	10	100.00%
6	0	0	0	0	0	20	0	20	100.00%
7	0	0	0	0	0	0	20	20	100.00%
Total	2	14	18	14	10	20	20	98	96.94%

1=Super 33+ (Cold Weather) 2=33 Electrical Tape 3=Super 88 4=Super 33+	5=Super 33+ (2001↓) 6=700 Commercial Grade 7=Temflex

The F value plot (Figure 8.1.6) produced after ANOVA indicated that the elements that aided discrimination between these black adhesives electrical tape brands were calcium and antimony.

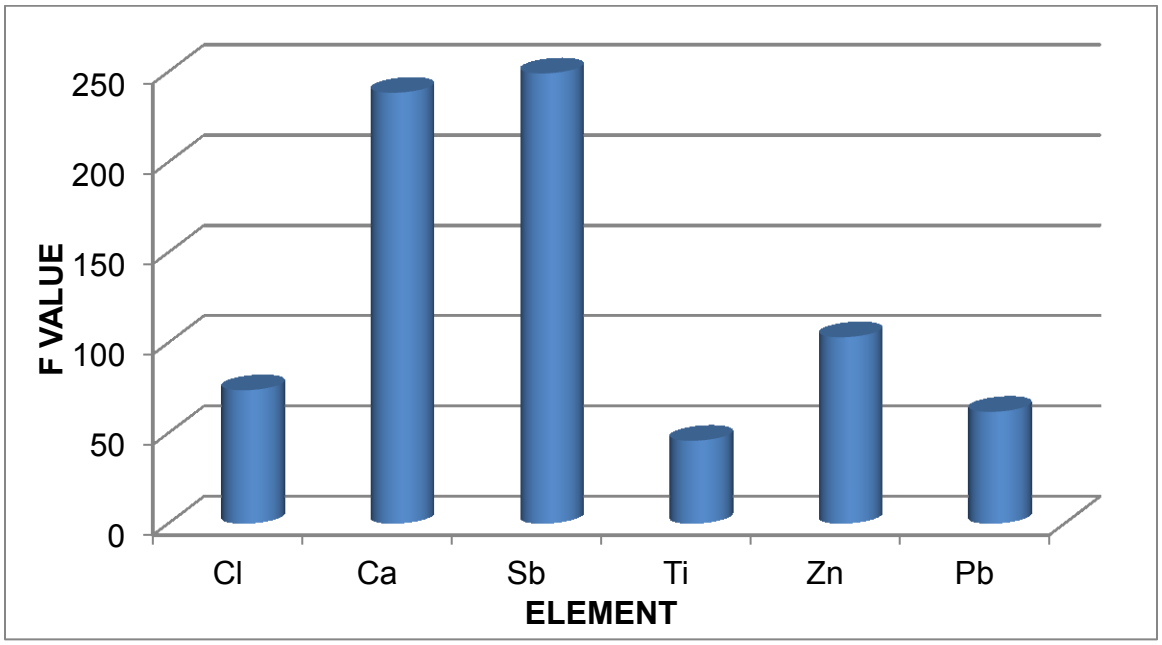


Figure 8.1.6 ANOVA F-values plot for black adhesives analyzed by XRF

8.2. XRF Colorless Adhesives

The list of samples with clear adhesive can be found in Table 8.2.1 which also includes their identifiers. Overall, there were forty-one samples in this group representing twenty-six brands.

Table 8.2.1 Colorless adhesive samples analyzed by XRF

Brand	Identifier
3M Tartan	3MA-C
ACE	ET1
All Temp	ET2
Auto Solutions	ET3
Bengal	ET4A-E
Champion	ET5A-D
Duck	ET6A-C
Duck 668 Pro	ET7
Electro tuff	ET8
Frost King (ET60)	ET9A-C
Frost King (ET60FR)	ET9D-E
Globe	ET10
GE	ET11
Intertape	ET12

Brand	Identifier
LePage's	ET13
Leviton	ET14
Manco	ET15
MIT	ET16
Nashua A-7	ET17
Power First	ET18
Permacel AW 8 ½	ET19
Permacel P-29	ET20
Radio Shack	ET21
Shurtape	ET22
Powerworks	ET23
Vanguard	ET24A-B
WUL	ET25

According to AHC these forty-one samples can be clustered into four groups as indicated by the truncation line in Figure 8.2.1. Right away it is noticed that sample 9 was split between two groups because these samples had different manufacture origins. Sample 6 also had samples from different origins and was split between two classes. The top star in the dendrogram indicates that confusion was seen between the samples of Champion Quality Tools and Duck. The bottom star indicates confusion between the samples of Shurtape and Powerworks. The central objects plot (Figure 8.2.2) indicated that Class 1 contained trace amounts of the other elements but mostly contained chlorine. Class 2 contained some calcium, aluminum, silicon, titanium, and lead. Class 3 contained a large amount of calcium and smaller amounts of titanium, iron, and lead. Class 4 contained a large amount of zinc and smaller amounts of calcium, silicon, and lead.

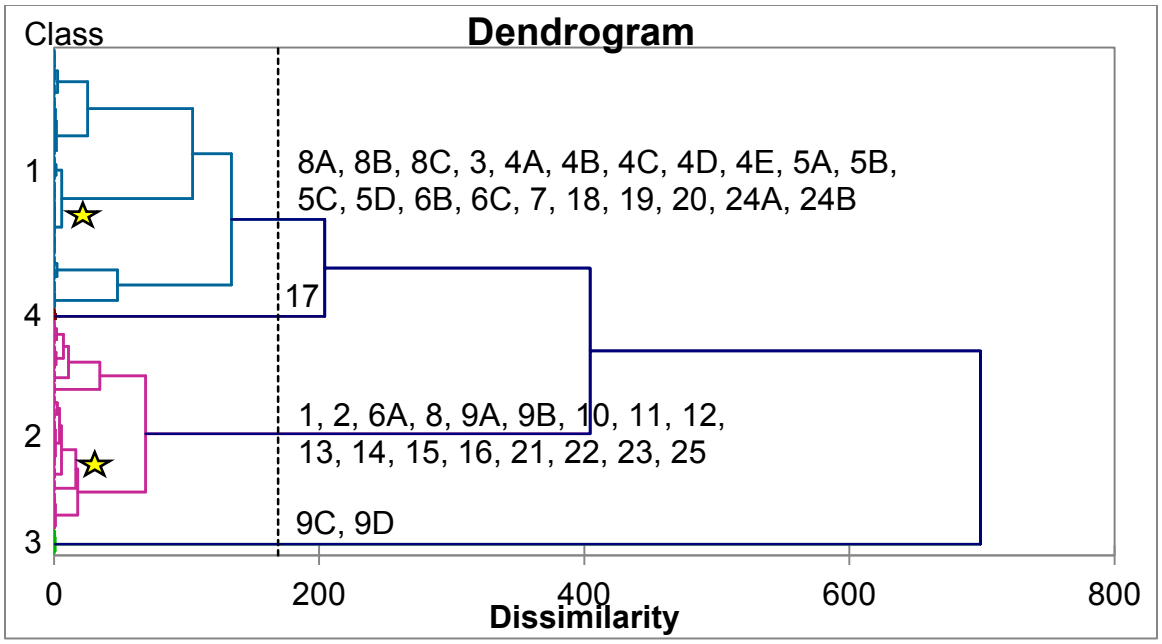


Figure 8.2.1 AHC dendrogram for colorless adhesives analyzed with XRF

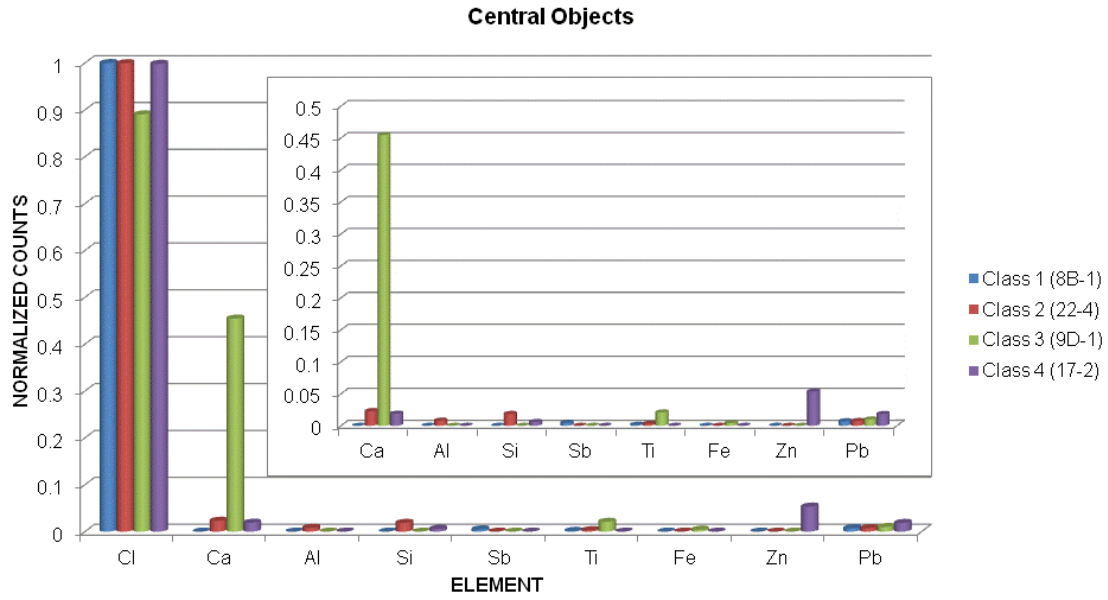
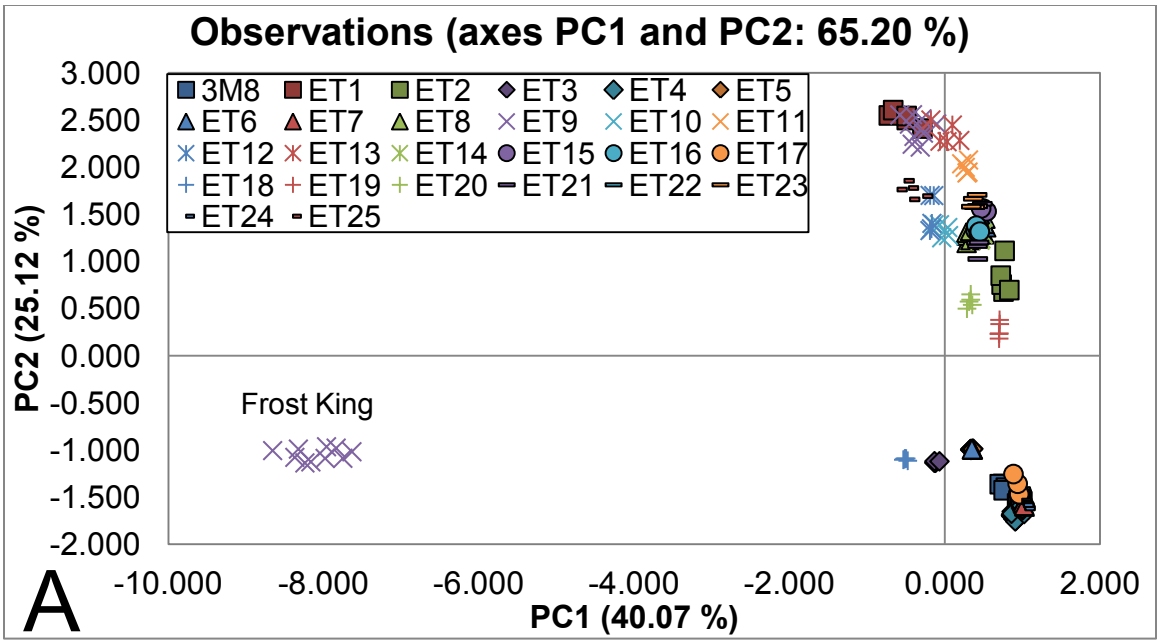


Figure 8.2.2 AHC central objects for colorless adhesives analyzed by XRF

PCA showed that the first two PCs captured 65.20% of the total variance while the first three PCs captured 79.56%. The observations plot (Figure 8.2.3) showed that better separation was seen when three dimensions are considered. Figure 8.2.3A, a plot of the first two PCs, showed that Sample 9 (Frost King ET60FR) was clearly distinguishable from the other samples. This is consistent with the results of AHC, where these samples were placed in a group by themselves. Otherwise, it was seen that the samples clustered around PC2 with a clear distinction between the larger dendrogram classes (Class 1 above PC1 and Class 2 below). Figure 8.2.3B is a plot of PC2 and PC3. A better separation was seen between the samples hovering around PC2. Now, Sample 17 (Nashua A-7) separated from large group of samples. Both brands of Permacel are also separated from the large group of samples.



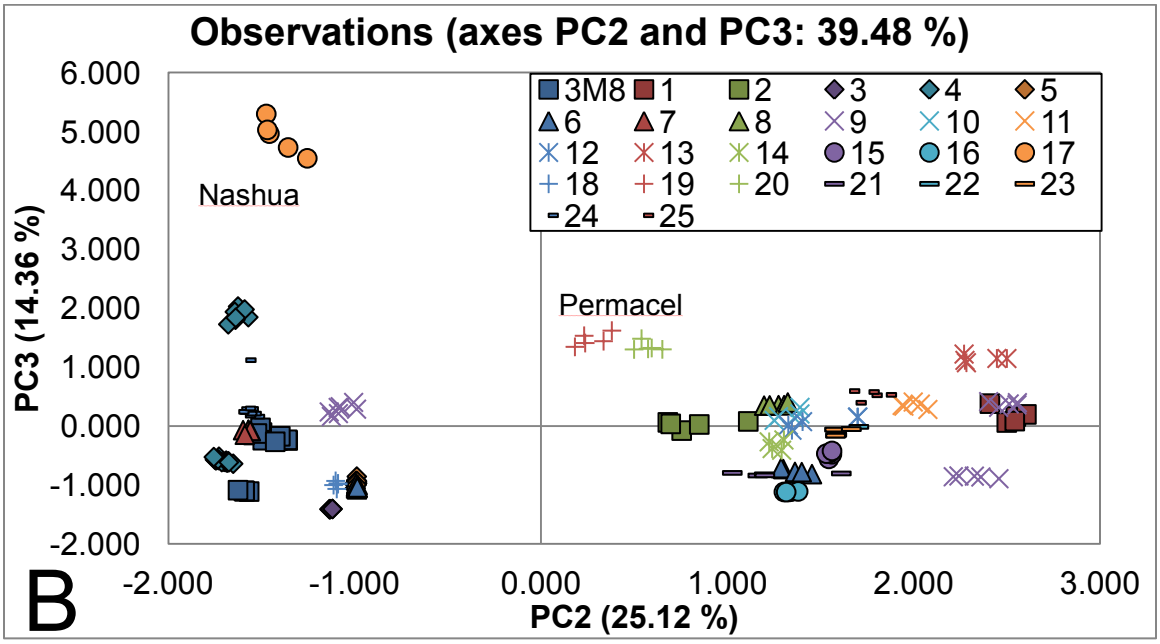


Figure 8.2.3 PCA observations plot for colorless adhesives analyzed by XRF

The factor loadings plot in Figure 8.2.4 indicated that PC1 showed negative correlations to calcium, titanium, and iron. PC1 showed a positive correlation to chlorine. PC2 showed positive correlations to aluminum and silicon. Lastly, PC3 showed positive correlations to zinc and lead. PC3 had no negative correlation with any element. Therefore, samples to the right on PC1 contain more chlorine or less calcium, titanium, and iron (e.g. in Figure 8.2.3A, the Frost King ET60FR samples contain more calcium, titanium, and iron than the other samples). Those samples to the right on PC2 contain more aluminum and silicon than those samples further to the left in Figure 8.2.3B. Lastly those samples further north on PC3 contain more zinc and lead than those samples further south (e.g. Nashua A-7 contains more zinc and lead together than the Tartan samples, which are the blue squared in Figure 8.2.3B).

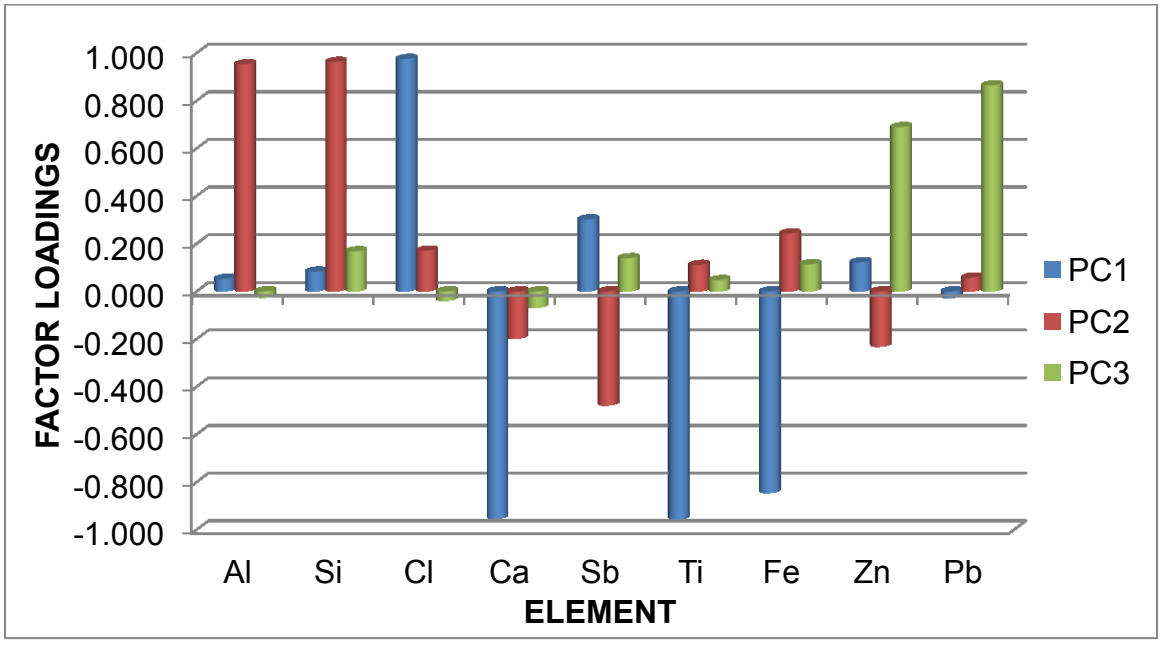


Figure 8.2.4 PCA factor loadings plot for colorless adhesives analyzed by XRF

DA showed a 100% classification accuracy when the samples were placed into groups according to AHC (data not shown). However, the accuracy decreased to 82.76% when each brand was assigned its own group. The observations plot in Figure 8.2.5 showed results similar to those in Figure 8.2.3B. Nashua A-7 was out on its own while the remaining brands formed two large clusters. The confusion matrix in Table 8.2.2 indicated that Tartan was misclassified as Duck 668 Pro. Duck was misclassified as Leviton and Manco. Samples of Duck 668 Pro were misclassified as Vanguard and vice versa. Samples from Frost King were misclassified as Globe and WUL. Samples from Leviton and Manco were misclassified as each other. Shurtape and Powerworks were completely misclassified as each other. The remaining brands were classified correctly 100% of the time.

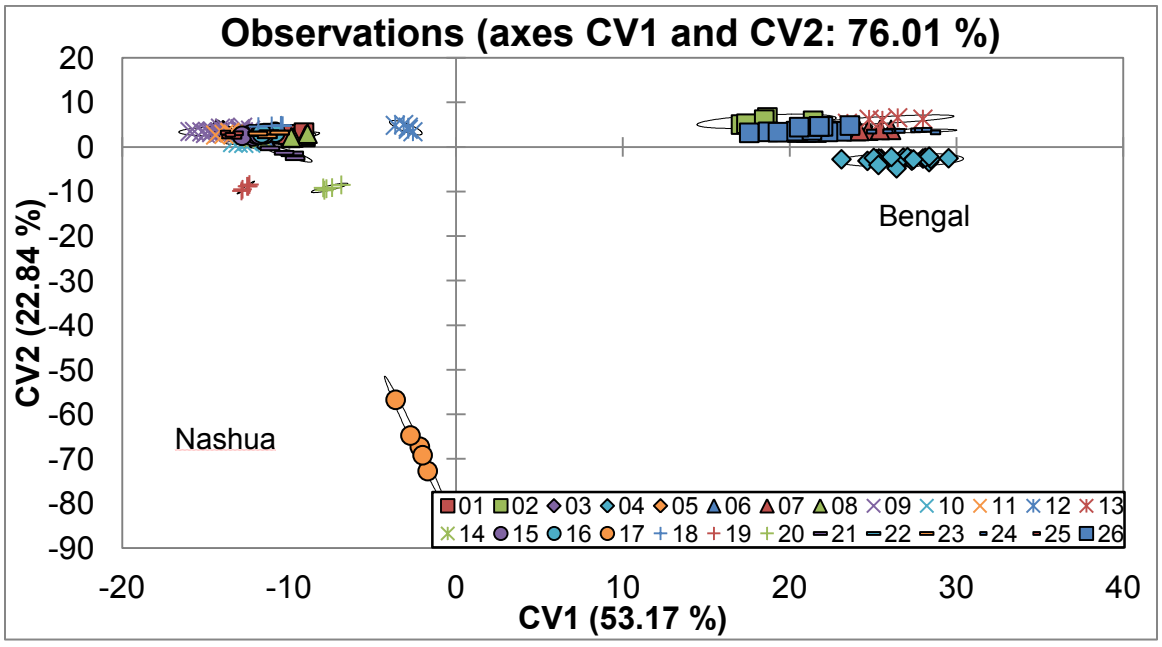


Figure 8.2.5 DA observations plot for colorless adhesives analyzed by XRF

Table 8.2.2 DA confusion matrix for colorless adhesives analyzed by XRF

Predicted / Actual	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total	% Correct
1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
3	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
4	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	100.00%
5	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	100.00%
6	0	0	0	0	10	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	15	0.00%
7	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	5	80.00%
8	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
9	0	0	0	0	0	0	0	0	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	20	85.00%
10	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
11	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
12	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
13	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5	100.00%
14	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	4	74.00%
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	5	80.00%
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5	100.00%
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	5	100.00%
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5	100.00%
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	100.00%
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5	100.00%
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	5	100.00%
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	5	0.00%
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	5	0.00%
24	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	10	80.00%
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	5	100.00%
26	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	15	86.67%
Total	5	5	5	24	30	0	8	5	17	6	5	5	5	8	6	5	5	5	5	5	5	5	5	9	7	13	203	82.76%

1=ACE 2=All Temp 3=AutoSolutions 4=Bengal 5=Champion	6=Duck 7=Duck 668 Pro 8=Electrotuff 9=Frost King 10=Globe	11=GE 12=Intertape 13=LePage's 14=Leviton	15=Manco 16=MIT 17=Nashua 18=Power First
19=Permacel AW 8½ 20=Permacel P-29 21=Radio Shack 22=Shurtape	23=Powerworks 24=Vanguard 25=WUL 26=Tartan		

Lastly, the F values plot produced from ANOVA (Figure 8.2.6) indicated that the most discriminating elements for these colorless adhesives electrical tape brands were antimony and zinc.

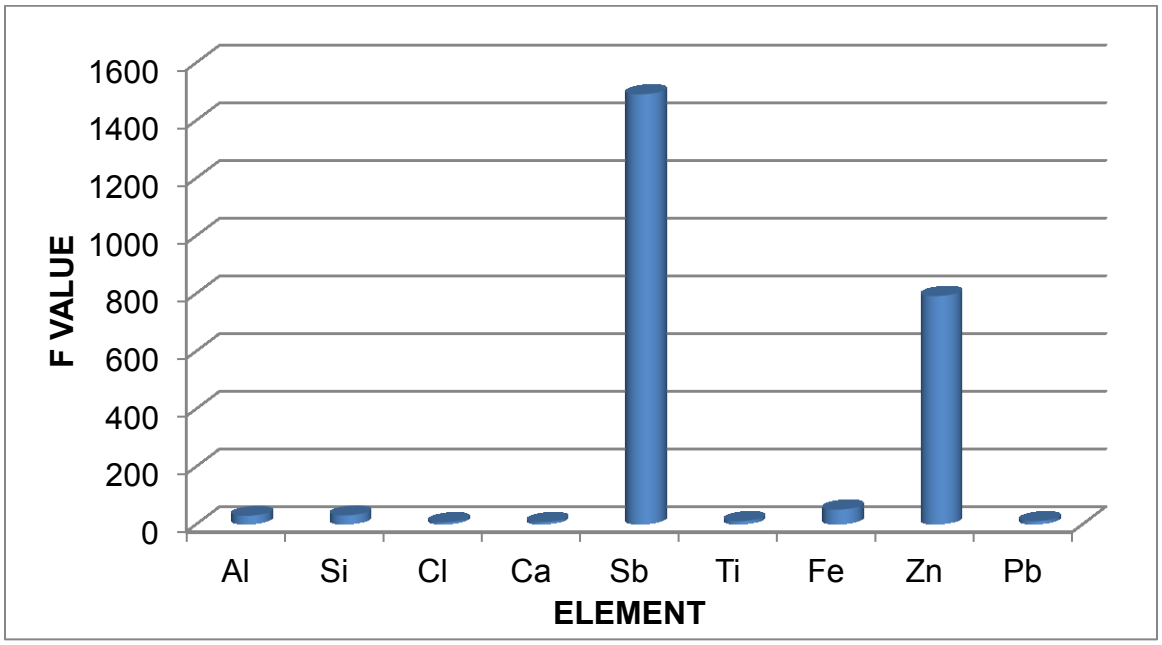


Figure 8.2.6 ANOVA F-values plot for colorless adhesives analyzed by XRF

8.3. DART/MS Positive Ion Mode Black Adhesives

The first part of the study was to perform multivariate statistical analysis on data collected in positive-ion mode. First, the rolls with black adhesive were analyzed. These samples are listed in Table 8.3.1. Ultimately, this group contained six brands of 3M electrical tapes.

Table 8.3.1 Black adhesive samples analyzed by DART/MS

Brand	Identifier	Brand	Identifier	Brand	Identifier
Super 88	BPT10	Super 33+	BPT26	Super 33+	BPT54
Super 33+	BPT11	Tartan	BPT27	88T	BPT61
Super 33+	BPT12	Tartan	BPT28	Super 88	BPT63
Super 33+	BPT13	Temflex	BPT29	Super 33+	BPT64
Super 33+	BPT15	Temflex	BPT30	700 Commercial	BPT65
Tartan	BPT16	Tartan	BPT34	Super 33+	BPT68
Super 88	BPT17	Tartan	BPT36	700 Commercial	BPT74
Super 33+	BPT18	Super 33+	BPT39	Super 33+	BPT76
Super 33+	BPT19	Tartan	BPT40	Super 33+	BPT77
Super 33+	BPT20	Super 33+	BPT41	Tartan	BPT78
Super 88	BPT23	Tartan	BPT43	700 Commercial	BPT79
Super 88	BPT24	Tartan	BPT44	Super 88	BPT80
Super 33+	BPT25	Tartan	BPT47	Super 33+	BPT83

AHC revealed three main groups within the data set. The first group contained both 3M Super 88 and 3M Super 33+ rolls of tape. The second group contained the 3M Tartan and 3M Temflex rolls. The third group contained the 3M 700 Commercial Grade rolls and a single roll of 3M Super 88. The dendrogram is shown in Figure 8.3.1. The dissimilarity between the two days was determined for each roll. The dissimilarity ranged from 0.000243 to 0.379. The representative spectra for each class are shown in the central objects plot (Figure 8.3.2). This showed that there were only small differences between the three classes.

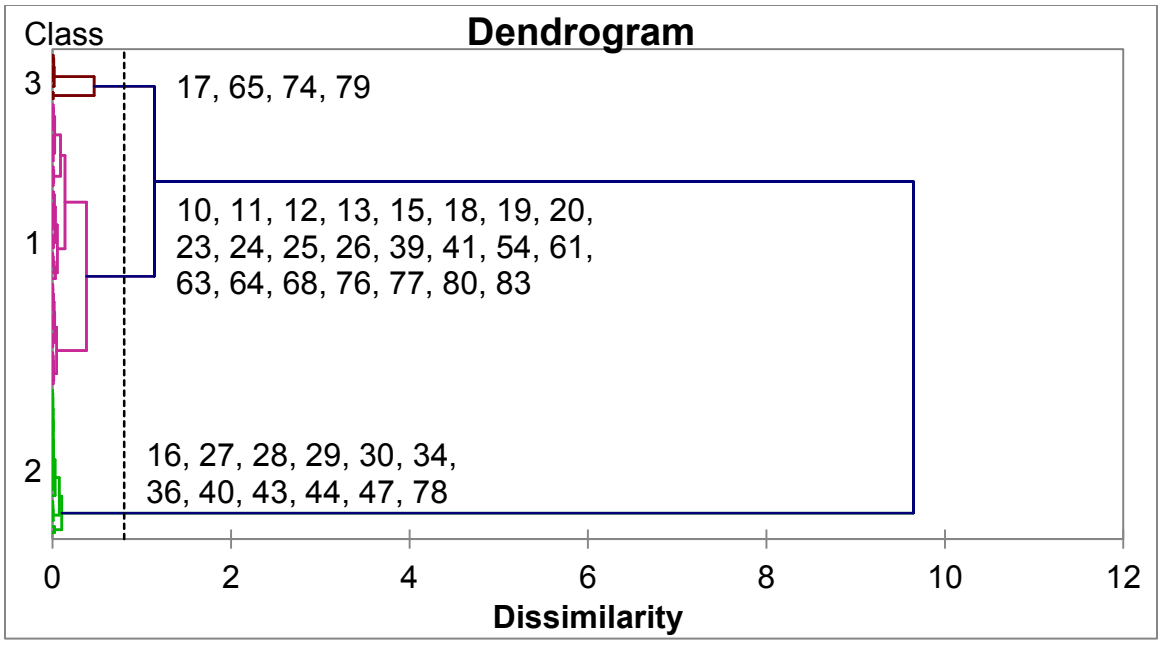


Figure 8.3.1 AHC dendrogram for black adhesives analyzed by DART/MS

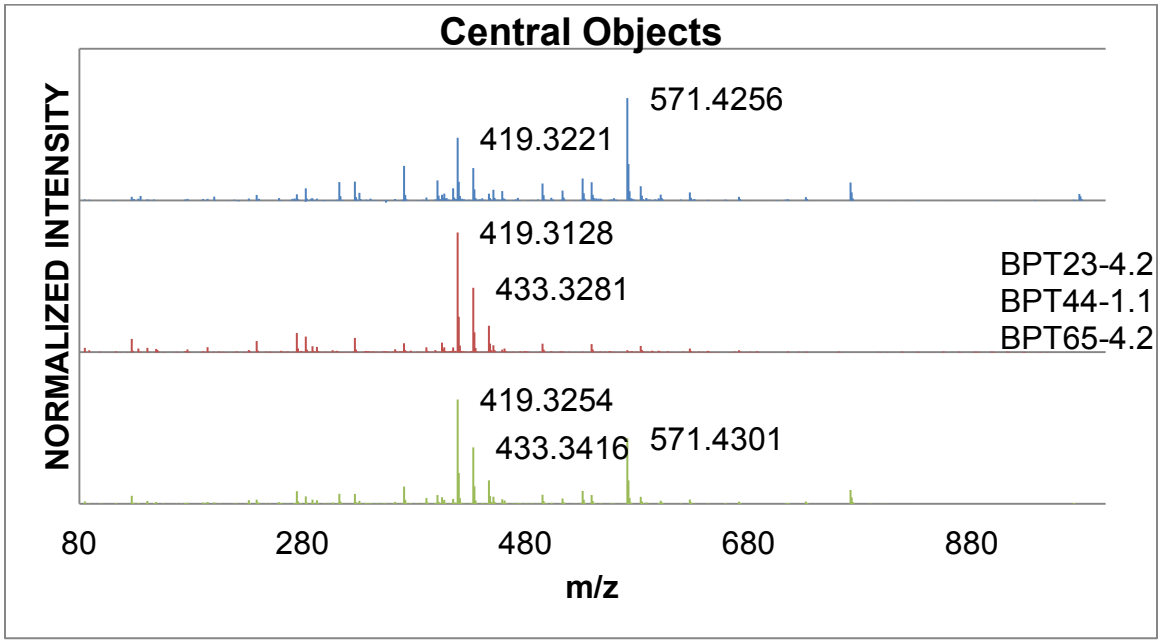


Figure 8.3.2 AHC central objects plot for black adhesives analyzed by DART/MS

PCA revealed through the observations plot that the three groups formed by AHC analysis are clearly distinguishable from one another. The total variance captured within the first two principal components was 52.26%. Figure 8.3.3 shows the observations plot coded according to sample. The observations plot showed that sample 17 (3M Super 88) was unique, for it is off on its own on PC2. Therefore, it was distinguishable from the other samples. There was a small distinction between the two groups at the bottom of the plot, consistent with the dendrogram results. It is hard to distinguish between the members of the respective groups. Even when the third dimension was considered, no additional information was obtained. Placement on the observations plot can be explained further by examining the factor loadings plot.

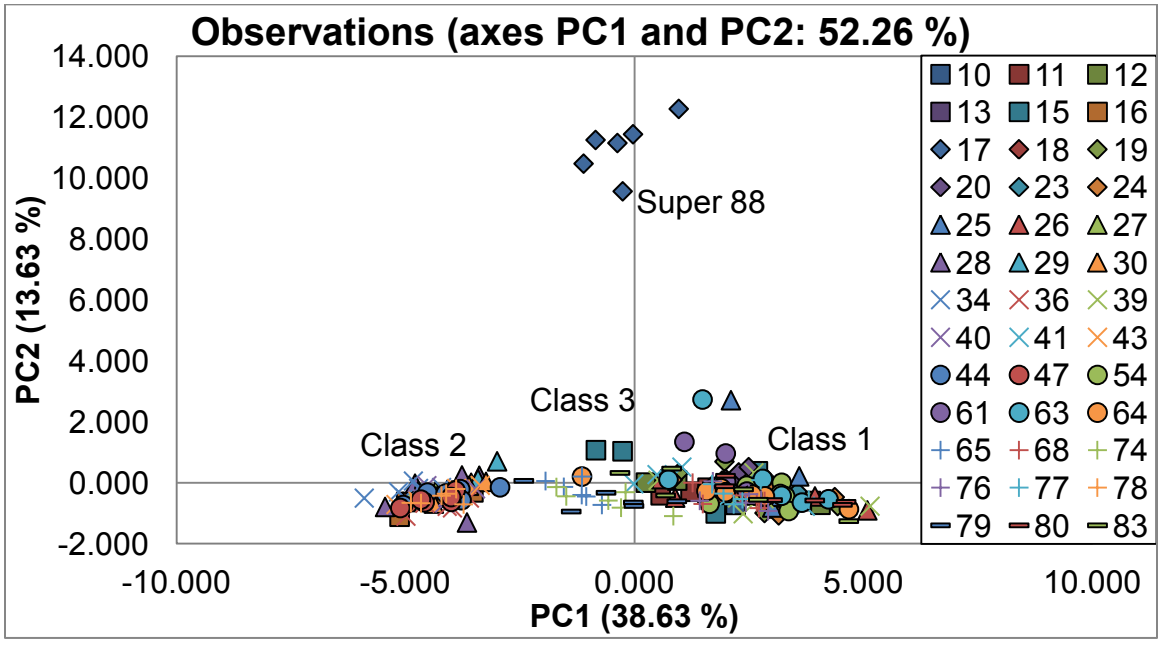


Figure 8.3.3 PCA observations plot for black adhesives analyzed by DART/MS

The factor loadings plot (Figure 8.3.4) indicated that PC1 had a positive correlation to m/z values of 313, 531, 571, 572, and 772. Table 8.3.2 summarizes the standards run under the same conditions. Based on these standards the m/z of 313 resulted from BBP. The other m/z values are currently

unknown. PC1 also showed a positive correlation to DOA with m/z values of 371, 372, and 373. PC1 showed a negative correlation to m/z values of 419, 420, 421, 433, 434, and 447. All of these m/z values are seen in the DNP standard; although, most are small. PC1 also showed a negative correlation to m/z values of 127 and 275. The sources of these m/z values are also currently unknown. PC2 showed no negative correlation to any m/z values but did show a positive correlation to m/z values of 299, 485, and 557. The m/z of 299 is seen in the DNP standard but the intensity was small. Otherwise, the sources of the other m/z values are unknown. Lastly, PC2 had a positive correlation to DOA with m/z values of 371, 372, and 373. Relating this back to the observations plot we know that Class 1 contained higher amounts of those compounds with a positive correlation to PC1 than Classes 2, 3, and 4. Class 2 contained higher amounts of the compounds with a negative correlation to PC1. Class 3 contained much larger amounts of those compounds with a positive correlation to PC2. Class 4 contained larger amounts than Class 1 but smaller amounts than Class 2 of the compounds with a negative correlation to PC1.

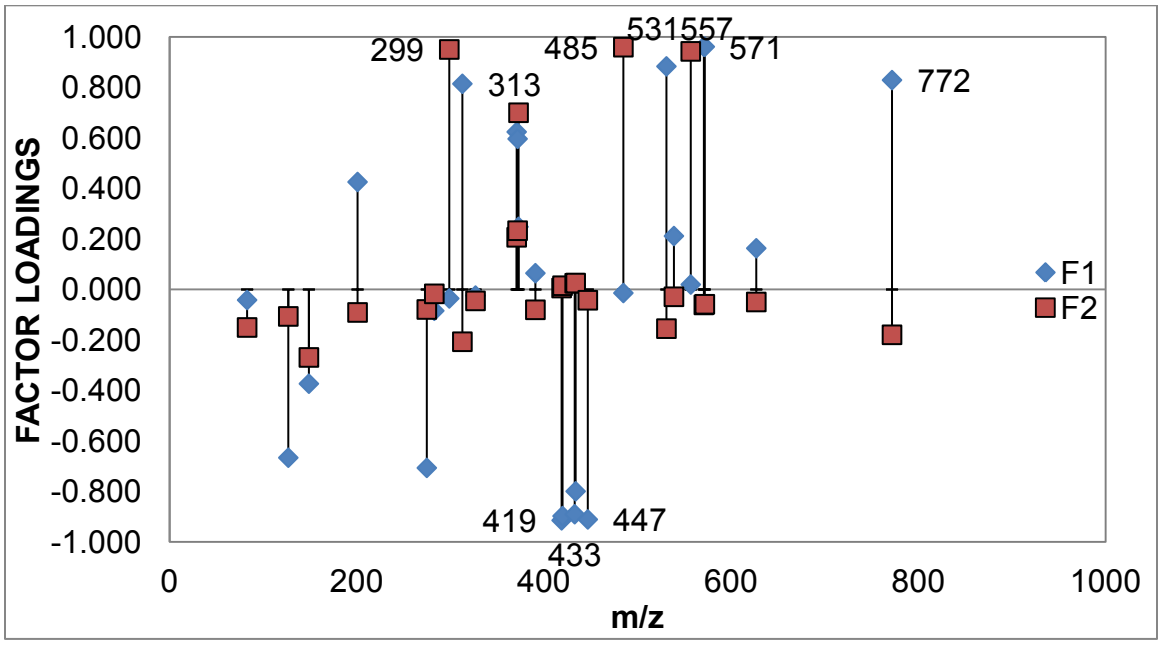


Figure 8.3.4 PCA factor loadings plot for black adhesives analyzed by DART/MS

Table 8.3.2 Plasticizer Standards analyzed by DART/MS

Name	Mass	Accurate Mass (M+H)
Dioctyl adipate	370.57	371.3225, *371.3187
Dioctyl phthalate	390.56	391.2900, *391.2809
Di-n-nonyl phthalate	418.61	419.3217
Trioctyl trimellitate	546.78	547.4056
Butyl benzyl phthalate	312.36	*313.1423
Dimethyl phthalate	194.18	*195.0637
Diethyl phthalate	222.24	*223.0944
Di-n-butyl phthalate	278.34	*279.1565
* Indicates analysis performed on the mixture		

DA was run several times for the tapes with black adhesive. The first, assigned the rolls to one of three classes according to the clustering from AHC. The overall classification accuracy was 99.12% with samples from Class 3 correctly classified 100% of the time (data not shown). A sample from Class 1 was incorrectly assigned to Class 3 and a sample from Class 2 was incorrectly assigned to Class 1. Overall, it was possible to discriminate between classes. Second, each roll was assigned to its own group resulting in thirty-nine groups. The observations plot in Figure 8.3.5 showed the same pattern as the observations plot from PCA. The overall accuracy was 28.07% with samples from rolls 17, 34, and 65 correctly classified 100% of the time. The remaining samples were classified as indicated in Table 8.3.3. The parentheses indicate how many samples were misclassified. In general, rolls of 3M Super 88 were constantly confused with 3M Super 33+ and other rolls of 3M Super 88. Likewise, rolls of 3M Super 33+ were constantly confused with 3M Super 88 and other rolls of 3M Super 33+. Rolls of 3M Tartan were constantly confused with 3M Temflex and other rolls of 3M Tartan. The same can be said for rolls of 3M Temflex. Rolls of 3M 700 Commercial Grade were only confused with other rolls of 3M 700 Commerical Grade. Lastly, each roll was assigned to a group

determined by Mehltretter *et al*.,[71] after visual and instrumental analysis. This resulted in eleven groups for the thirty-nine samples. The overall accuracy was 67.54% (data not shown); there were more misclassifications between the different products. Only one group, group 27 had all samples correctly classified. Groups containing 3M Temflex and 3M Tartan were once again confused with other 3M Temflex and 3M Tartan samples. The groups containing 3M Super 88 and 3M Super 33+ were not confused with groups containing 3M Tartan and 3M Temflex but were confused with groups containing 3M Super 33+, 3M Super 88, and 3M 700 Commercial Grade. The groups containing 3M 700 Commercial Grade were confused with groups containing 3M Tartan and 3M Temflex, which was not seen before.

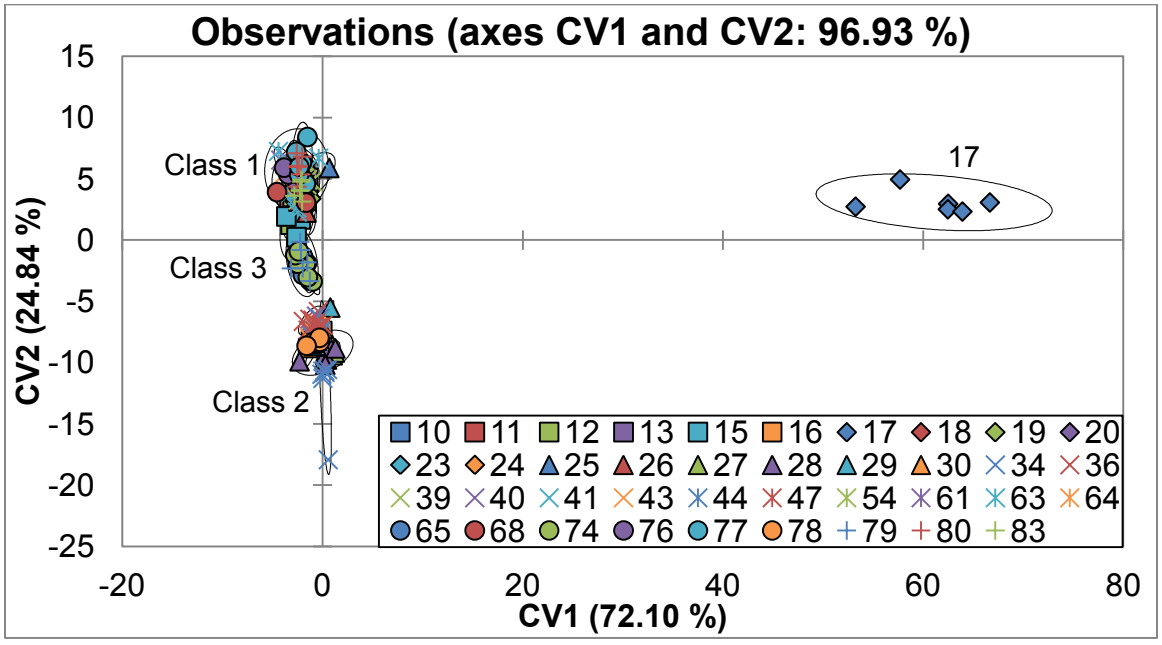


Figure 8.3.5 DA observations plot for black adhesives analyzed by DART/MS

Table 8.3.3 DA confusion matrix for black adhesives analyzed by DART/MS

Actual	Predicted						Total	% Correct
10	13 (1)	18(1)	68 (1)				3/5	40.00%
11	10 (1)	13 (1)					2/5	60.00%
12	13 (1)	15 (1)	18(1)	26 (1)	61 (1)		5/5	0.00%
13	10 (1)	12 (1)	18(1)	20 (1)	64 (1)		5/5	0.00%
15	11 (1)	12 (1)	18(1)	65 (1)			4/5	20.00%
16	29 (1)	30 (1)	43 (1)	44 (2)			5/6	16.67%
18	12 (1)	19 (1)					2/6	66.67%
19	23 (1)	25 (1)	41 (1)	54 (1)	80 (1)		5/6	16.67
20	19 (1)	25 (2)					3/6	50.00%
23	18(1)	20 (1)	39 (1)	41 (1)	54 (1)		5/6	16.67%
24	15 (1)	20 (1)	23 (2)	41 (1)			5/5	0.00%
25	19 (1)	20 (2)	23 (1)	24 (1)	63 (1)		6/6	0.00%
26	10 (1)	12 (2)	15 (1)	18(2)			6/6	0.00%
27	28 (3)	47 (1)					4/6	33.33%
28	27 (6)						6/6	0.00%
29	16 (1)	30 (1)	43 (2)	78 (1)			5/6	16.67%
30	16 (1)	29 (2)	43 (2)	44 (1)			6/6	0.00%
36	16 (1)	43 (1)	47 (3)				5/6	16.67%
39	10 (1)	23 (1)	41 (1)	54 (1)	64 (1)		5/6	16.67%
40	29 (1)	30 (1)	43 (3)	64 (1)			6/6	0.00%
41	15 (1)	19 (1)	54 (2)				4/6	33.33%
43	29 (1)	30 (2)	47 (2)				5/6	16.67%
44	16 (1)	29 (1)	47 (2)	78 (1)			5/6	16.67%
47	36 (3)	43 (2)	44 (1)				6/6	0.00%
54	13 (1)	18(1)	25 (1)	39 (2)	64 (1)		6/6	0.00%
61	13 (1)	23 (1)	26 (1)	41 (1)	63 (2)		6/6	0.00%
63	25 (1)	26 (2)	61 (2)	64 (1)			6/6	0.00%
64	12 (1)	41 (1)	80 (1)				3/6	50.00%
68	10 (1)	64 (2)	83 (2)				5/6	16.67%
74	65 (1)	79 (1)					2/6	66.67%
76	20 (1)	77 (1)	83 (1)				3/6	50.00%
77	76 (1)	83 (1)					2/6	66.67%
78	16 (1)	30 (1)					2/6	66.67%
79	65 (1)	74 (4)					5/6	16.67%
80	23 (1)	64 (2)	76 (1)	77 (1)			5/6	16.67%
83	20 (1)	26 (1)	41 (1)	64 (1)	76 (1)	80 (1)	6/6	0.00%
Super 88 10, 17, 23, 24, 61, 63, 80		Super 33+ 11-13, 15, 18, 19, 20, 25, 26, 39, 41, 54, 64, 68, 76, 77, 83		Tartan, Temflex 16, 27-30, 34, 36, 40, 43, 44, 47, 78		700 Commercial 65, 74, 79		28.07%

Finally, the F-value plot in Figure 8.3.6 indicated that the intensity of m/z values of 485 contributed the most to the discrimination between the black electrical tape samples. Unfortunately, the source of the m/z of 485 is currently unknown. Other m/z values contributing to the discrimination of samples included 299, 447, 557, 571, and 572. As was mentioned above, the m/z values of 299 and 447 were seen in the standard of DNP but at low intensities. The remaining m/z values are unknown as well.

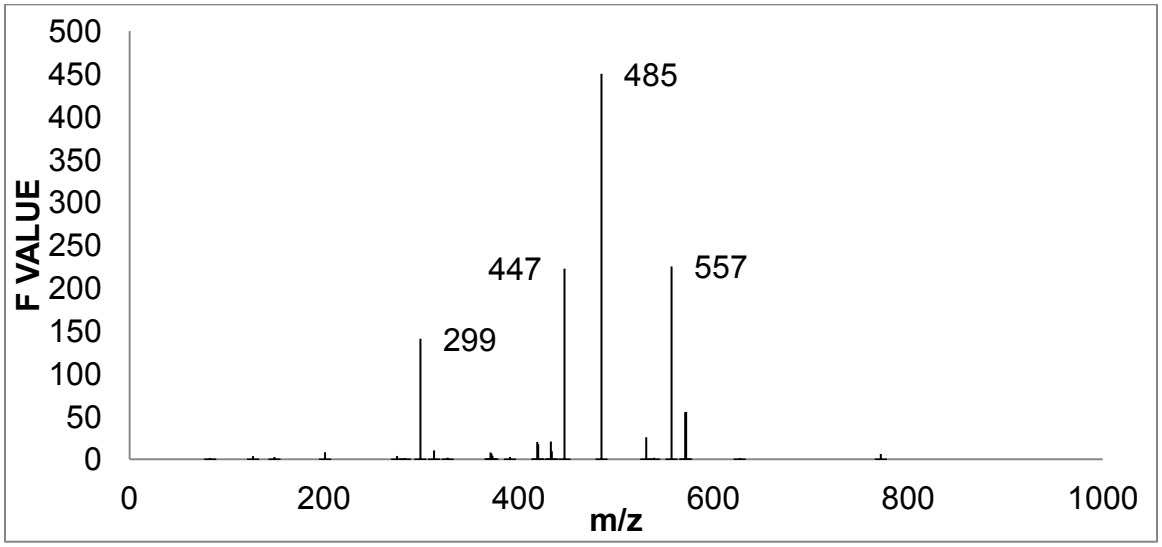


Figure 8.3.6 ANOVA F-value plot for black adhesives analyzed by DART/MS

8.4. DART/MS Positive Ion Mode Tinted Adhesives

After the black adhesive rolls were analyzed, the clear-brown-tinted adhesive tapes were analyzed in the same manner. However, since there were only three rolls of tape (33, 53, and 62) that were part of this group the reliability of the analysis is questionable. With that said the three rolls were clearly distinguishable from one another in all three statistical techniques (AHC, PCA, DA). There was a positive correlation between PC1 and m/z values of 275, 313, 401, 419, and 420. PC1 showed a negative correlation to m/z values of 199, 217, 231, 297, 345, 389, 402, and 417. PC2 showed positive correlations to m/z values of 149, 279, 391, 392, 547, and 548. Lastly, the m/z values that discriminated the samples were 419 and 420 both due to DOP (no data shown).

8.5. DART/MS Positive Ion Mode Colorless Adhesives

Lastly, the colorless adhesive electrical tapes were analyzed. There were a total of forty-seven rolls with clear-colorless adhesive (listed in Table 8.5.1).

Table 8.5.1 Colorless adhesive samples analyzed by DART/MS

Brand	Identifier
Marcy	BPT1
Advance	BPT2
WorkSaver	BPT3
Tesa Tape	BPT4
Tape It	BPT5
Qualpack	BPT6
Marcy	BPT7
Manco	BPT8
Archer	BPT9
Frost King ET60	BPT14
Manco	BPT21
Manco	BPT22
Regal	BPT31
GE	BPT32
Frost King ET60	BPT35
National	BPT37

Brand	Identifier
Manco	BPT38
National	BPT42
Calterm	BPT45
Manco	BPT46
Tape It	BPT48
Tape It	BPT49
Gen Electric	BPT50
National	BPT51
National	BPT53
Manco	BPT55
Victor Auto	BPT56
United	BPT57
Frost King ET60	BPT58
Tuff Hand	BPT59
Tuff Hand	BPT60
L. G.	BPT66

Brand	Identifier
Manco	BPT67
Tartan	BPT69
Tyco	BPT70
Qualpack	BPT71
Nitto Denko	BPT72
Frost King ET60FR	BPT73
Ace	BPT81
Ace	BPT82
Tartan	BPT84
Frost King ET60FR	BPT85
Duck	BPT86
Nitti Denko	BPT87
Frost King ET60FR	BPT88
Power Pro	BPT89
Duck	BPT90

According to AHC, the forty-seven rolls can be separated into six groups, as shown in the dendrogram (Figure 8.5.1). From this analysis, Qualpack, Tuff, and 3M Tartan General Use were placed into classes by themselves. The remaining brands were divided into three classes. All samples within the large clusters were not clustering according to brand as before with the black adhesives. A plot of the central objects illustrates features that separate the classes (refer to Figure 8.5.2). Class 6 was unique because it was the only class with a different highest intensity m/z value.

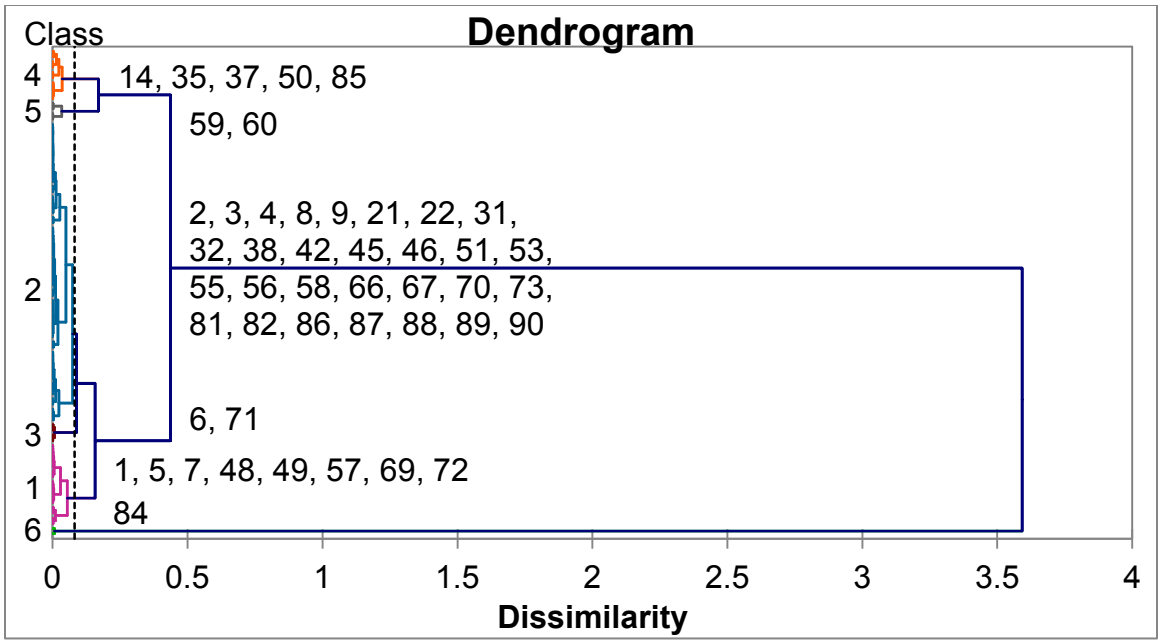


Figure 8.5.1 AHC dendrogram for colorless adhesives analyzed by DART/MS

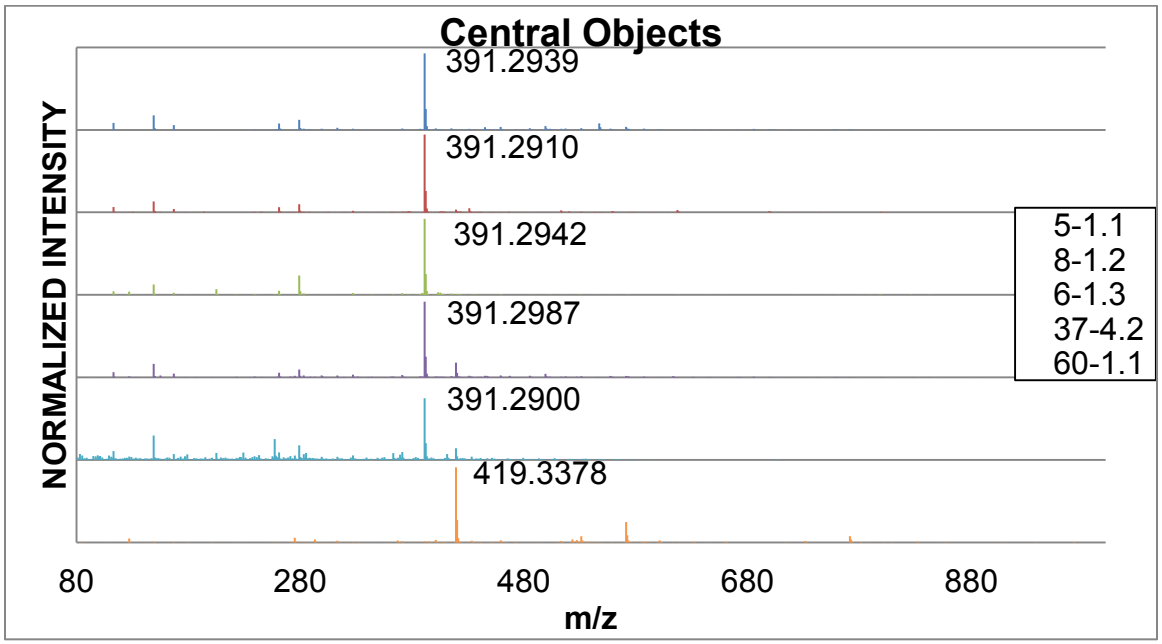


Figure 8.5.2 AHC central objects plot for colorless adhesives analyzed by DART/MS

An observations plot (Figure 8.5.3) produced from PCA showed that the first two PCs captured 39.58% of the total variance, while the first three PCs captured 50.07% of the total variance. The observation plot is coded according to AHC class placement and showed that Class 5, containing both Tuff rolls, and Class 6, containing 3M Tartan General uses, were distinguishable from the other classes. Class 1, containing Marcy, Tape It, United, Nitto Denko, and Tartan rolls were not distinguishable from one another and slightly distinguishable from Classes 2, 3, and 4 when the third dimension was considered. Class 4, containing some rolls of Frost King, National, and General Electric, was slightly distinguishable from the Class 1. However, Classes 2 and 3 were not distinguishable from each other as shown by complete overlapping. Even when the third dimension is considered Class 3 is not distinguishable from Class 2.

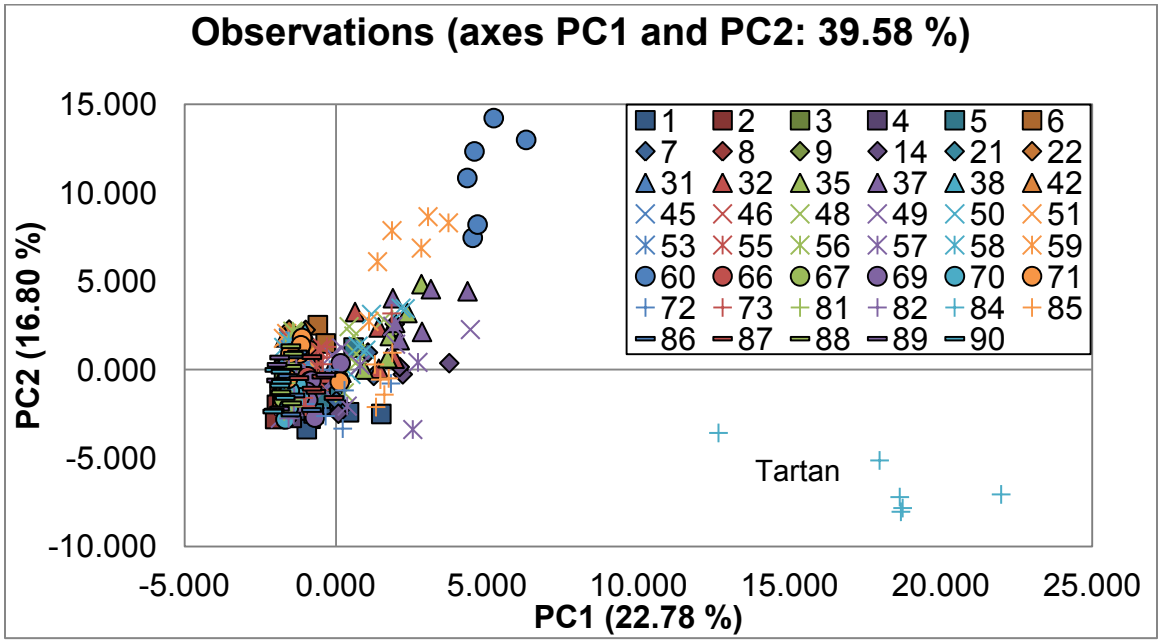


Figure 8.5.3 PCA observations plot for colorless adhesives analyzed by DART/MS

The factor loadings plot for this data set is shown in Figure 8.5.4. There was a negative correlation between PC1 and m/z values of 391, 392, and 393 which correspond to DOP, a common plasticizer. There was a positive correlation between PC1 and m/z values of 419, 420, and 421 which corresponds to DNP, another plasticizer. There was also a strong positive correlation with m/z values of 571 and 572 with origins currently unknown. PC2 had a positive correlation to the compounds whose m/z values were less than 500.

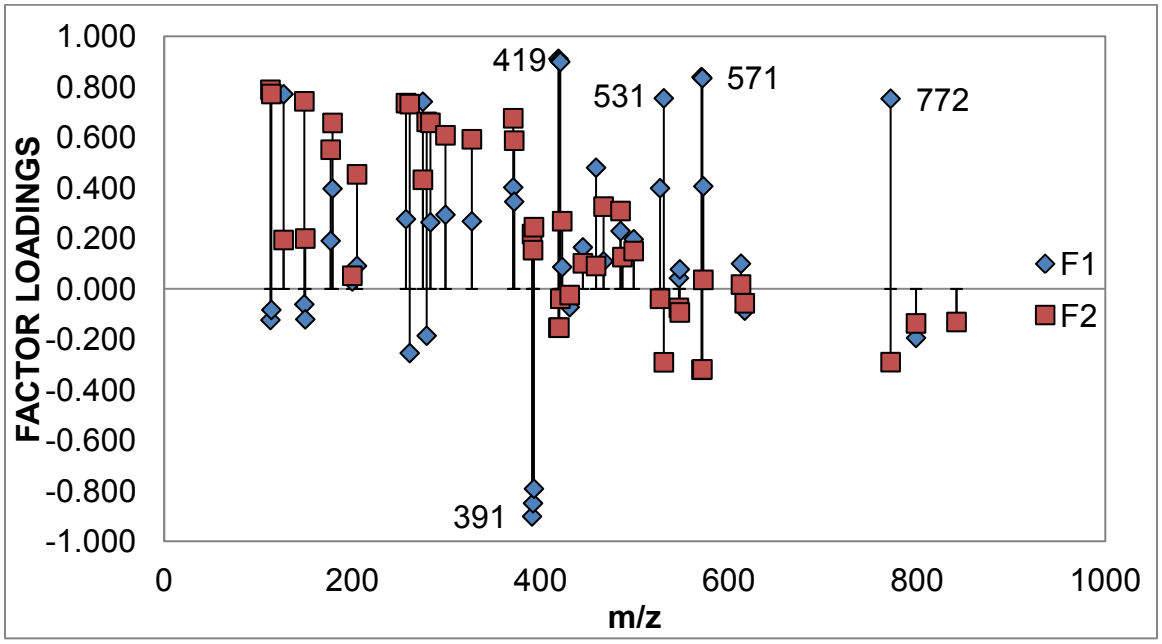


Figure 8.5.4 PCA factor loadings plot for colorless adhesives analyzed by DART/MS

DA was performed several times, each time with a different classification scheme. First, the rolls of tape were assigned to a class designated by the clustering from AHC. Therefore, they were placed into one of six classes. This resulted in an overall classification accuracy of 97.07% (data not shown). Samples from Classes 3, 4, and 6 were correctly classified 100%. Samples from Classes 1 and 5 were incorrectly classified as Class 2 samples. A few samples from Class 2 were misclassified as Class 4. The observations plot showed that

Classes 3 and 8 are clearly distinguishable while Classes 5 and 6 are slightly distinguishable. However, Classes 1, 2, 4, and 7 are overlapped. The second analysis assigned each rolls of tape to its own class. Therefore, there were forty-seven groups. The observations plot, in Figure 8.5.5, showed that rolls 6, 71, and 84 were clearly distinct from the other rolls however; rolls 6 and 71 were not distinct consistent with being the same brand. Rolls 59, 60, and 85 were slightly distinct from the other rolls; however, roll 59 overlapped with some replicates from the larger cluster. Rolls 14, 35, 37, and 50 which were clustered together in the dendrogram were once again clustered in the observations plot and are separated from the remaining cluster of rolls. Rolls clustered in Classes 1, 2, 4, and 6 in the dendrogram were completely overlapped in the observations plot. Overall, 50.92% of the samples were correctly classified. Table 8.5.2 gives more detail on the classification of the samples. Rolls 2, 57, and 85 were classified correctly 100% of the time. The same confusion pattern was seen throughout every analysis. If they clustered together in the dendrogram or observations plots, they were confused most often with one another. The last analysis assigned each roll to a previously determined group (as published in Mehltretter *et al.*).[71] Therefore, each roll was assigned to one of twenty-five groups (data not shown). The same trend was seen in this observation plot as in Figure 8.5.5. Groups 6, 14, 15, 19, and 39 which contained rolls 85, 6, 71, 59 and 60, and 84, respectively, were unique from the other rolls. Group 8 which contained rolls 14, 35, 37, and 50 was slightly separated from the large cluster of rolls. The remaining groups were clustered on top of one another. Overall, the classification accuracy was 70.70%; with samples from groups 6, 18, and 39 classified correctly 100% of the time. No samples from groups 2, 3, 10, 11, 13, 16, 17, and 24 were classified correctly. The majority of the confusion occurred with group 9. However, only two samples from group 9 were confused with any other group.

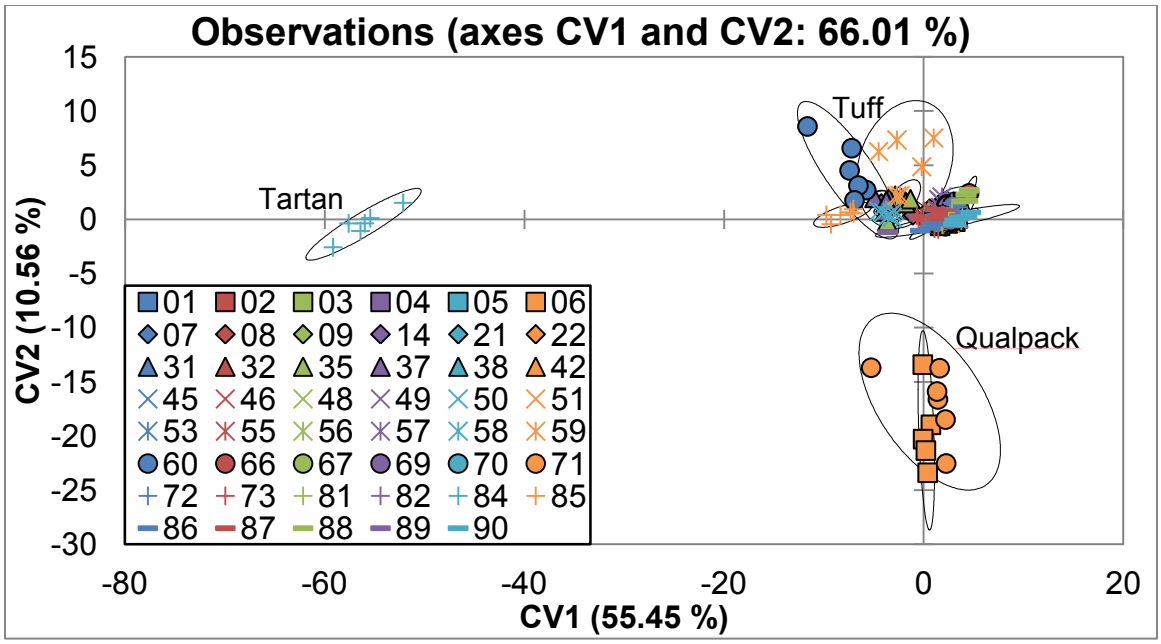


Figure 8.5.5 DA observations plot for colorless adhesives analyzed by DART/MS

Table 8.5.2 DA confusion matrix for colorless adhesives analyzed by DART/MS

Actual	Predicted					Total	% Correct
1	2 (1)	5 (2)	7 (2)			5/6	16.67%
3	2 (1)	89 (1)				2/5	40.00%
4	4 (1)	81 (2)				3/5	60.00%
5	7 (1)					1/5	80.00%
6	71 (1)					1/5	80.00%
7	1 (1)	5 (2)				3/5	40.00%
8	22 (1)	81 (1)				2/5	60.00%
9	31 (1)	81 (1)				2/5	60.00%
14	35 (1)					1/5	80.00%
21	8 (2)	22 (2)	66 (1)			5/6	16.67%
22	8 (1)	21 (2)	38 (2)	46 (1)		6/6	0.00%
31	9 (1)	51 (1)	81 (1)			3/6	50.00%
32	14 (1)	87 (1)				2/6	66.67%
35	37 (1)					1/6	83.33%
37	35 (4)					4/6	33.33%
38	8 (2)	45 (1)				3/6	50.00%
42	9 (1)	51 (1)	56 (3)	90 (1)		6/6	0.00%
45	55 (3)	58 (1)				4/6	33.33%
46	31 (1)					1/5	80.00%
48	49 (2)	86 (1)				3/6	50.00%

Table 8.5.2 Continued.

Actual	Predicted					Total	% Correct
49	5 (1)	7 (1)	48 (2)			4/6	33.33%
50	35 (2)					2/6	66.67%
51	9 (1)	53 (2)				3/6	50.00%
53	2 (2)	51 (4)				6/6	0.00%
55	45 (3)	58 (2)	86 (1)			6/6	0.00%
56	9 (1)	31 (1)	51 (2)	70 (1)		5/6	16.67%
58	51 (1)	81 (2)	89 (1)			4/6	33.33%
59	60 (1)					1/6	83.33%
60	59 (1)					1/6	83.33%
66	21 (1)	67 (2)				3/6	50.00%
67	8 (1)	21 (2)	70 (1)			4/6	33.33%
69	5 (1)	67 (1)				2/6	66.67%
70	67 (1)	82 (2)	86 (2)			5/6	16.67%
71	6 (1)					1/6	83.33%
72	5 (1)					1/6	83.33%
73	4 (1)	57 (1)				2/6	66.67%
81	2 (1)	4 (1)	31 (1)	42 (1)	89 (1)	5/6	16.67%
82	4 (1)	31 (2)	81 (2)	86 (1)		6/6	0.00%
86	4 (1)	56 (2)	81 (2)	82 (1)		6/6	0.00%
87	82 (1)					1/6	83.33%
88	89 (1)					1/6	83.33%
89	2 (2)	51 (1)	58 (1)	86 (1)		5/6	16.67%
90	2 (1)	89 (1)				2/6	66.67%
							50.92%

Lastly, the F-value plot revealed that m/z values of 391, 419, 420, and 421 contributed the most to the discrimination between samples (refer to Figure 8.5.6). The m/z value of 391 corresponds to DNP and the m/z values of 419-421 correspond to DOP, both plasticizers.

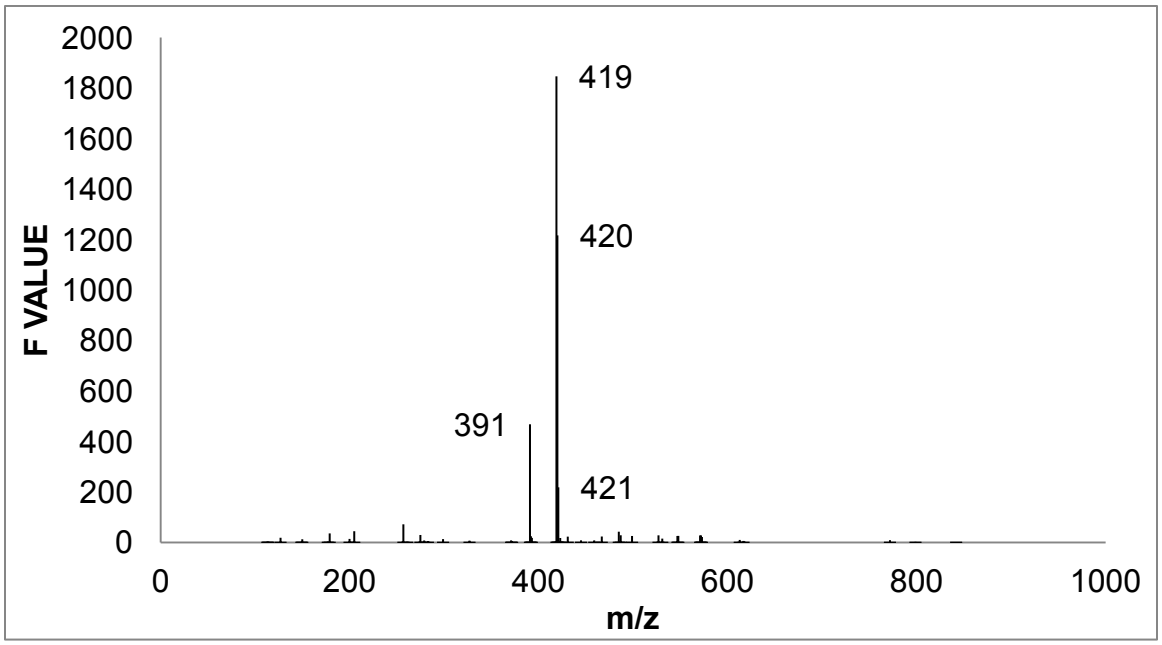


Figure 8.5.6 ANOVA F-value plot for colorless adhesives analyzed by DART/MS

8.6. DART/MS Negative Ion Mode

The second part of the study was to perform multivariate statistical analysis on the data obtained in negative-ion mode. However, after analysis it was determined that this technique was lacking reproducibility. Due to this, there will not be a full discussion of the chemometric results. Therefore, tape analysis by DART-MS should be performed in positive-ion mode.

CHAPTER 9. CONCLUSIONS

The XRF analysis of tapes with black adhesives indicated that the different brands can be discriminated from one another except for Super 33+ (Cold Weather) and Super 88. This was supported by the results of PCA and DA analysis. It is also noted that a clear difference was seen between the manufacture years of Super 33+ based on the presence of lead. However, the elements that aided in the discrimination between brands were calcium and antimony.

The tapes with the colorless adhesive largely placed in two different classes based on the presence of aluminosilicate filler. Nashua A-7 and Frost King ET60FR can be considered unique because there were placed into their own group during AHC analysis and were readily distinguishable during PCA and DA analysis. According to DA, most of the samples were distinguishable although, this was not supported by the results of AHC or PCA analysis. The elements that contributed the most to the discrimination of the samples were antimony and zinc.

Overall, XRF was shown to be an appropriate technique for the analysis of electrical tape. However, this technique did not provide any additional discrimination than seen within SEM/EDS analysis. The colorless adhesives formed two major groups; those with and those without aluminosilicate filler. Within those two groups brand discrimination was difficult.

The chemometric analysis of DART/MS analysis of black adhesive tapes revealed that a roll of Super 88 and the rolls of 700 Commercial Grade were readily distinguishable from the other tapes. Super 88 and Super 33+ are not distinguishable from one another as indicated in AHC, PCA, and DA. Tartan and

Temflex were also not distinguishable from one another using any technique. Lastly, the most discriminating m/z values were 299, 447, 557, 571, and 572 with most origins currently unknown.

The results of chemometric analysis on the colorless adhesive tapes showed that Qualpack, Tuff Hand Tools, and a roll of Tartan were unique in PCA, and DA analysis. Otherwise, the samples were placed into three classes. Individual brand discrimination was shown to be difficult. The most discriminating m/z values for this data set were 391, 419, 420, and 421 which were due to the plasticizers Di-n-nonyl phthalate and Dioctly phthalate. Overall, DART/MS does not seem to be an appropriate technique based on the results described above.

CHAPTER 10. RECOMMENDATIONS

The first recommendation involves color coordinates. It is preferred that the spectra be obtained using reflectance. Therefore, it is suggested that reflectance spectra be acquired using the Uv-vis MSP for the dyes previously analyzed and perform the color coordinates again. Another option is to perform the color coordinate calculation on the absorbance spectra rather than the transmittance. This is acceptable and usually published as complementary color coordinates. This may improve the accuracy of the coordinates.

The second recommendation involves the tape samples analyzed by DART/MS. It is recommended that chemometric techniques be employed on the other techniques (i.e. FT-IR, SEM/EDS, and Py-GC). These spectra and chromatograms are usually stored upon analysis. Therefore, they can be exported and sent to IUPUI so that the chemometric techniques can be performed. Then it can be officially determined whether DART/MS provided additional separation between the electrical tape brands and should be employed during the tape analysis scheme.

LIST OF REFERENCES

LIST OF REFERENCES

1. Morgan, S. L.; Bartick, E. G., Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, Blackledge, R. D., Ed. John Wiley & Sons, Inc.: New York, 2007; pp 333-374.

2. National, C. R., *Strengthening Forensic Science in the United States: a path forward*. National Academies Press: Washington, DC, 2009; p 352.

3. William Daubert, et ux., etc., et al., petitioners v. Merrell Dow Pharmaceuticals, Inc. U.S. Supreme Court: 1993; Vol. 509 US 579, 589.

4. Johnson, R.; Wichern, D., *Applied Multivariate Statistical Analysis*. Prentice Hall: Englewood Cliffs, NJ, 1982; p 594.

5. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc.: New York, 1998; p 348.

6. Brereton, R. G., *Chemometrics for Pattern Recognition*. John Wiley & Sons, Inc.: Chichester, England, 2009; p 504.

7. Kramer, R., *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker, Inc: New York, 1998; p 203.

8. Brereton, R. G., *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Ltd.: Chichester, England, 2003; p 489.

9. Brereton, R. G., *Chemomentrics: Applications of Mathematics and Statistics to Laboratory Systems*. Ellis Horwood Limited: Chichester, England, 1990; p 307.

10. Sharaf, M. A.; Illman, D. L.; Kowalski, B. R., *Chemometrics*. John Wiley & Sons, Inc.: 1986; p 332.

11. Manly, B. F. J., *Multivariate Statistical Methods: A primer*. 3rd ed.; Chapman & Hall/CRC: 2005; p 214.

12. StatSoft, I., Electronic Statistics Textbook. StatSoft: Tulsa, OK, 2011. http://www.statsoft.com/textbook/ (accessed January 2012).

13. Gallo, J. M.; Almirall, J. R., Elemental analysis of white cotton fiber evidence using solution ICP-MS and laser ablation ICP-MS (LA-ICP-MS). *Forensic Science International* **2009,** *190* (1-3), 52-57.

14. Hida, M.; Sato, H.; Sugawara, H.; Mitsui, T., Classification of counterfiet coins using multivariate analysis with X-ray diffraction and X-ray fluroescence methods. *Forensic Science International* **2001,** *115*, 129-134.

15. Goodpaster, J. V.; Sturdevant, A. B.; Kristen L. Andrews; Brun-Conti, L., Identification and Comparison of Electrical Tapes Using Instrumental and Statistical Techniques: I. Microscopic Surface Texture and Elemental Composition. *Journal of Forensic Sciences* **2007,** *52* (3), 610-629.

16. Goodpaster, J. V.; Sturdevant, A. B.; andrews, K. L.; Briley, E. M.; Brun-Conti, L., Identification and Comparison of Electrical Tapes Using Instrumental and Statistical Techniques: II. Organic Composition of the Tape Backing and Adhesives. *Journal of Forensic Sciences* **2009,** *54* (2), 328-338.

17. Barrett, J. A.; Siegel, J. A.; Goodpaster, J. V., Forensic Discrimination of Dyed Hair Color: II. Multivariate Statistical Analysis. *Journal of Forensic Sciences* **2011,** *56* (1), 95-101.

18. Klemenc, S., In common batch searching of illicit heroin samples - evaluation of data by chemometrics methods. *Forensic Science International* **2000,** *115* (1-2), 43-52.

19. Mat Desa, W. N. S.; Daeid, N. N.; Ismail, D.; Savage, K., Application of Unsupervised Chemometric Analysis and Self-organizing Feature Map (SOFM) for the Classification of Lighter Fuels. *Analytical Chemistry* **2010,** *82* (15), 6395-6400.

20. Hibbard, R.; Goodpaster, J. V.; Evans, M. R., Factors Affecting the Forensic Examination of Automotive Lubricating Oils. *Journal of Forensic Sciences* **2011,** *56* (3), 741-753.

21. Muehlethaler, C.; Massonnet, G.; Esseiva, P., The application of chemometrics on Infrared and Raman spectra as a tool for the forensic analysis of paints. *Forensic Science International* **2011,** *209* (1-3), 173-182.

22. Thanasoulias, N. C.; Parisis, N. A.; Evmiridis, N. P., Multivariate chemomentrics for the forensic determination of blue ball-point pen pinks based on their Vis spectra. *Forensic Science International* **2003,** *138* (1-3), 75-84.

23. Chalmers, J. M.; Everall, N. J., FTIR, FT-Raman and chemometrics:applications to the analysis and characterisation of polymers. *Analytical Chemistry* **1996,** *15* (1), 18-25.

24. Thanasoulias, N. C.; Piliouris, E. T.; Kotti, M.-S. E.; Evmiridis, N. P., Application of multivariate chemometrics in forensic soil discrimination based on the UV-Vis spectrum of the acid fraction of humus. *Forensic Science International* **2002,** *130* (2-3), 73-82.

25. Chazottes, V.; Brocard, C.; Peyrot, B., Particle size of soils under simulated scene of crime conditions: the interest of multivariate analyses. *Forensic Science International* **2004,** *140*, 159-166.

26. Grimm, L.; Bryant, F. B.; Yarnold, P. R., *Reading and Understanding Multivariate Statistics*. American Psychological Association: Washington, DC, 1995; p 373.

27. Sandercock, P. M. L.; Pasquier, E. D., Chemical fingerprinting of unevaporated automotive gasoline samples. *Forensic Science International* **2003,** *134* (1), 1-10.

28. Sandercock, P. M. L.; Pasquier, E. D., Chemical fingerprinting of gasoline: 2. Comparison of unevaporated and evaporated automotive gasoline samples. *Forensic Science International* **2004,** *140* (1), 43-59.

29. Mahoney, C. M.; Gillen, G.; Fahey, A. J., Characterization of gunpowder samples using time-of-flight secondary ion mass spectrometry (TOF-SIMS). *Forensic Science International* **2006,** *158* (1), 39-51.

30. May, C. D.; Watling, J., The Development of Analytical and Interpretational Protocols to Facilitate the Provenance Establishment of Polycarbonate Headlamp Lens Material. *Journal of Forensic Sciences* **2011,** *56* (S1), S47-S57.

31. Turner, D. A.; Goodpaster, J. V., Comparing the Effects of Weathering and Microbial Degradation on Gasoline Using Principal Components Analysis. *Journal of Forensic Sciences* **2012,** *57* (1), 64-69.

32. Monfreda, M.; Gregori, A., Differentiation of Unevaporated Gasoline Samples According to Their Brands, by SPME–GC–MS and Multivariate Statistical Analysis. *Journal of Forensic Sciences* **2011,** *56* (2), 372-380.

33. Sarkissian, G., The Analysis of Tire Rubber Traces Collected After Braking Incidents Using Pyrolysis-Gas Chromatography/Mass Spectrometry. *Journal of Forensic Sciences* **2007,** *52* (5), 1050-1056.

34. Miller, J. C.; J. N, M., *Statistics for Analytical Chemistry*. 2nd ed.; Ellis Horwood: Chichester, England, 1988; p 227.

35. Anderson, R. L., *Practical Statistics for Analytical Chemists*. Van Nostrand Reinhold: New York, 1987; p 316.

36. McGaw, E. A.; Szymanski, D. W.; Smith, R. W., Determination of Trace Elemental Concentrations in Document Papers for Forensic Comparison Using Inductively Coupled Plasma-Mass Spectrometry. *Journal of Forensic Sciences* **2009,** *54* (5), 1163-1170.

37. Garvin, E. J.; Koons, R. D., Evaluation of Match Criteria Used for the Comparison of Refractive Index of Glass Fragments. *Journal of Forensic Sciences* **2011,** *56* (2), 491-500.

38. Udey, R. N.; Hunter, B. C.; Smith, R. W., Differentiation of Bullet Type Based on the Analysis of Gunshot Residue Using Inductively Coupled Plasma Mass Spectrometry. *Journal of Forensic Sciences* **2011,** *56* (5), 1268-1276.

39. Goodpaster, J. V., Discrimination of Dyed Cotton Fibers Based on UV-visible Microspectrophotometry and Multivariate Statistical Analysis. Midwest Forensic Resource Center: Indiana University Purdue University Indianapolis, 2010; Vol. 55,000.

40. David, S. K.; Pailthorpe, M. T., Classification of Textile Fibres: Production, Structure, and Properties. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 1-31.

41. Wiggins, K. G., Thin Layer Chromatographic Analysis of Fibre Dyes. In *Forensic Examination of Textile Fibres*, Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 291-310.

42. Adolf, F. P.; Dunlop, J., Microspectrophotmetry/Colour Measurement. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 251-290.

43. Grieve, M.; Dunlop, J.; Haddock, P., An Assessment of the Value of Blue, Red, and Black Cotton Fibers as Target Fibers in Forensic Science Investigations. *Journal of Forensic Sciences* **1988,** *33* (6), 1332-1344.

44. Grieve, M.; Dunlop, J.; Haddock, P., An Investigation of Known Blue, Red, and Black Dyes Used in teh Coloration of Cotton Fibers. *Journal of Forensic Sciences* **1990,** *35* (2), 301-315.

45. Paterson, M. D.; Cook, R., The Production of Colour Coordinates From Microgram Quantities of Textile Fibres. Part 1. *Forensic Science International* **1980,** *15*, 249-258.

46. Laing, D. K.; Dudley, R. J.; Isaacs, M. D. J., Colorimetric Measurements on Small Paint Fragments using Microspectrophotometry. *Forensic Science International* **1980,** *16*, 159-171.

47. Locke, J.; Cousins, D. R.; Russell, L. W.; Jenkins, C. M.; Wilkinson, J. M., A Data Collection of Vehicle Topcoat Colour. 1. Instrumentation for Colour Measurements. *Forensic Science International* **1987,** *34*, 131-142.

48. Laing, D. K.; Hartshorne, A. W.; Harwood, R. J., Colour Measurements on Single Textile Fibres. *Forensic Science International* **1986,** *30*, 65-77.

49. Hartshorne, A. W.; Laing, D. K., The Definition of Colour for Single Textile Fibres by Microspectrophotometry. *Forensic Science International* **1987,** *34*, 107-129.

50. Hartshorne, A. W.; Laing, D. K., Microspectrfluorimetry of Fluorescent Dyes and Brighteners on Single Textiles Fibres: Part 2 - Colour Measurements. *Forensic Science International* **1991,** *51*, 221-237.

51. Palenik, S. J., Microscopical Examination of Fibres. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 153-178.

52. Challinor, J. M.; Roux, C., Instrumental Methods Used in Fibre Examination. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 223-250.

53. Kirkbride, K. P.; Tungol, M. W., Infrared Microspectroscopy of Fibres. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 179-222.

54. Griffin, R.; Speers, J.; Robertson, J., Other Methods of Colour Analysis. In *Forensic Examination of Fibres*, 2 ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, 1999; pp 311-342.

55. SWGMAT, Forensic Fiber Examination Guidelines. *Forensic Science Communications* **1999,** *1* (1).

56. Smith, J. M., Forensic Examination of Pressure Sensitive Tape. In *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, Blackledge, R. D., Ed. John Wiley & Sons, Inc.: 2007.

57. SWGMAT, Guideline for the Forensic Examination of Pressure-Sensitive Tapes. *Forensic Science Communications* **2008,** *10* (4).

58. Jenkins, C. M., *X-ray Fluorescence Spectrometry*. John Wiley and Sons, Inc.: 1988; Vol. 99, p 175.

59. Jenkins, C. M., *X-ray Fluorescence Spectrometry*. 2 ed.; John Wiley and Sons, Inc.: 1999; Vol. 152, p 207.

60. Bertin, E. P., *Principles and Practice of X-ray Spectrometric Analysis*. Plenum Press: New York, 1970; p 679.

61. Zieba-Palus, J.; Borusiewicz, R.; Kunicki, M., PRAXIS-combined u-Raman and u-XRF spectrometers in the examination of forensic samples. *Forensic Science International* **2008,** *175*, 1-10.

62. Hida, M.; Mitsui, T.; Minami, Y., Forensic investigation of counterfeit coins. *Forensic Science International* **1997,** *89*, 21-26.

63. Trzcinska, B. M., Classification of Black Powder Toners on the Basis of Integrated Analytical Information Provided by Fourier Transform Infrared Spectrometry and X-Ray Fluorescence Spectrometry. *Journal of Forensic Sciences* **2006,** *51* (4), 919-924.

64. Flynn, J.; Stoilovic, M.; Lennard, C.; Prior, I.; Kobus, H., Evaluation of X-ray microfluorescence spectrometry for the elemental analysis of firearm discharge residues. *Forensic Science International* **1998,** *97* (1), 21-36.

65. Berendes, A.; Neimke, D.; Schumacher, R.; Barth, M., A Versatile Technique for the Investigation of Gunshot Residue Patterns on Fabrics and Other Surfaces: m-XRF. *Journal of Forensic Sciences* **2006,** *51* (5), 1085-1090.

66. Worley, C. G.; Wiltshire, S. S.; Miller, T. C.; Havrilla, G. J.; Majidi, V., Detection of Visible and Latent Fingerprints Using Micro-X-ray Fluorescence Elemental Imaging. *Journal of Forensic Sciences* **2006,** *51* (1), 57-63.

67. Laramee, J. A.; Cody, R. B.; Nilles, J. M.; Durst, H. D., Forensic Application of DART^{TM} (Direct Analysis in Real Time) Mass Spectrometry. In *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, Blackledge, R. D., Ed. John Wiley & Sons, Inc.: 2007.

68. Cody, R. B.; Laramee, J. A.; Nilles, J. M.; Durst, H. D., Direct Analysis in Real Time (DART^{tm}) Mass Spectrometry. *JOEL News* **2005,** *40* (1), 8-12.

69. Cody, R. B.; Laramee, J. A.; Durst, H. D., Versatile New Ion Source for the Analysis of Materials in Open Air Under Ambient Conditions. *Anal Chem* **2005,** *77* (8), 2297-2302.

70. Mehltretter, A. H.; Bradley, M. J.; Wright, D. M., Analysis and Discrimination of Electrical Tapes: Part I. Adhesives. *Journal of Forensic Sciences* **2011,** *56* (1), 82-94.

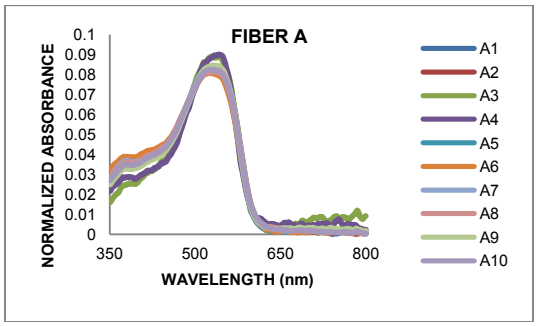
71. Mehltretter, A. H.; Bradley, M. J.; Wright, D. M., Analysis and Discrimination of Electrical Tapes: Part II. Backings. *Journal of Forensic Sciences* **2011**.

APPENDICES

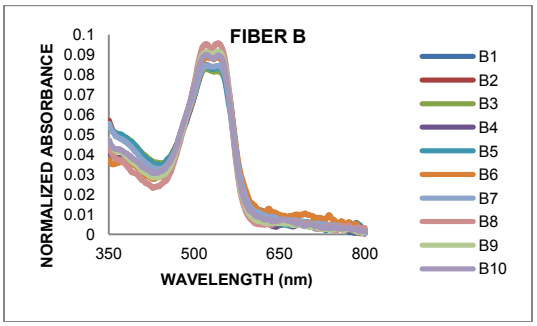
Appendix A. Fiber Dyes

A.1.Testfabrics, Inc. Dyes Analyzed at IUPUI

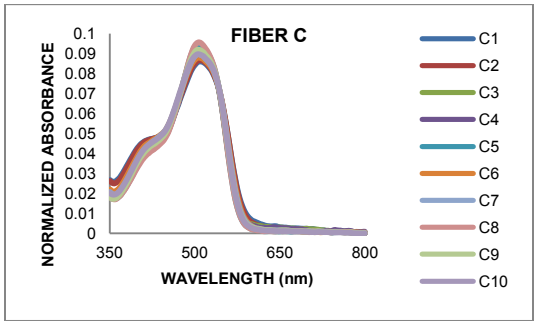
A.1.1. Training Set



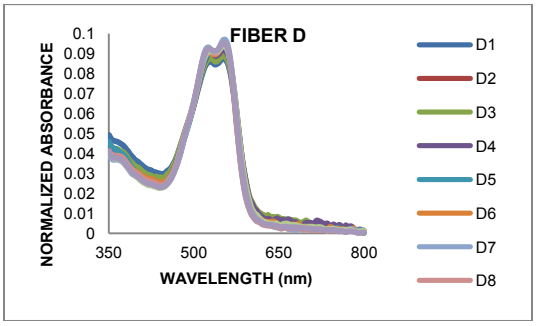
Averaged Spectra for Fiber A



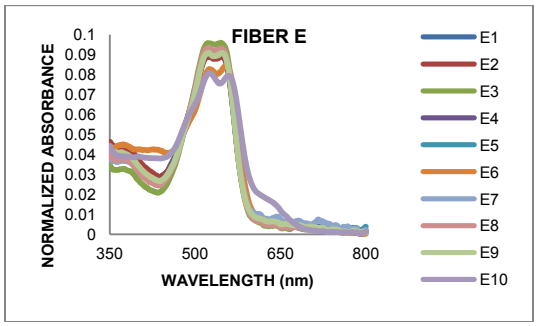
Averaged Spectra for Fiber B



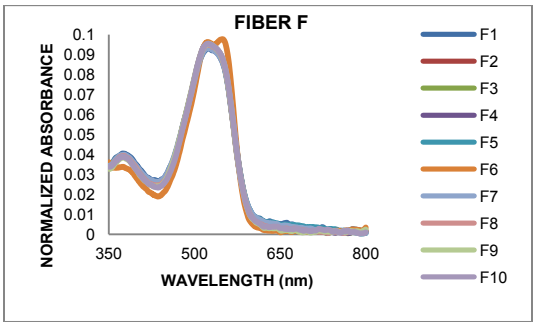
Averaged Spectra for Fiber C



Averaged Spectra for Fiber D

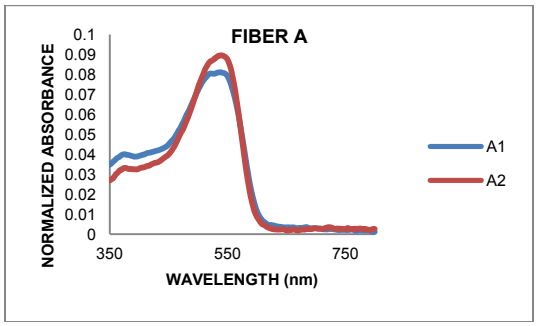


Averaged Spectra for Fiber E

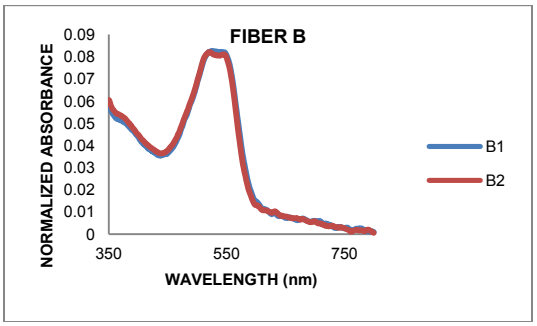


Averaged Spectra for Fiber F

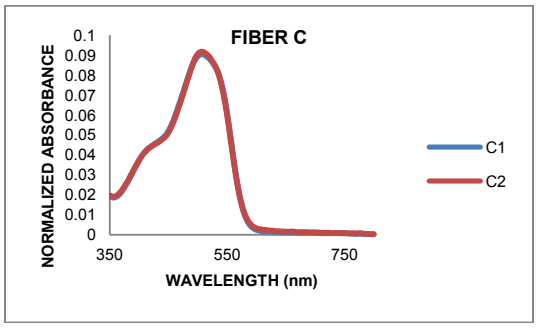
A.1.2. External Validation



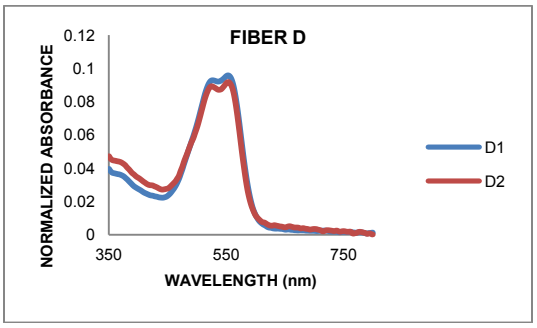
Averaged Spectra for EV Fiber A



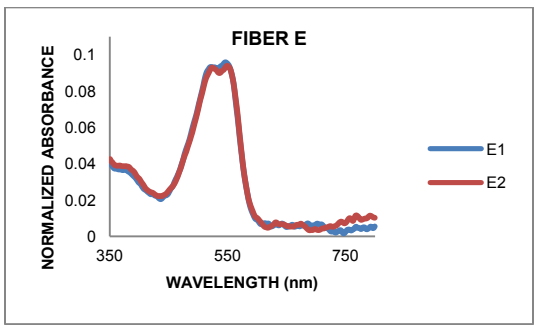
Averaged Spectra for EV Fiber B



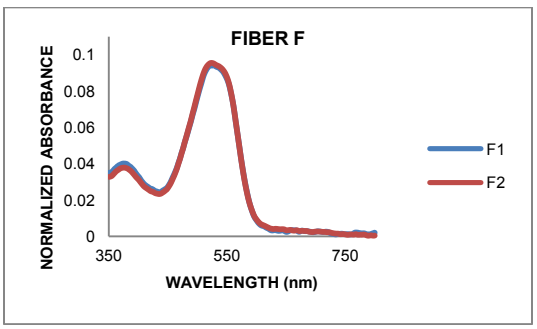
Averaged Spectra for EV Fiber C



Averaged Spectra for EV Fiber D



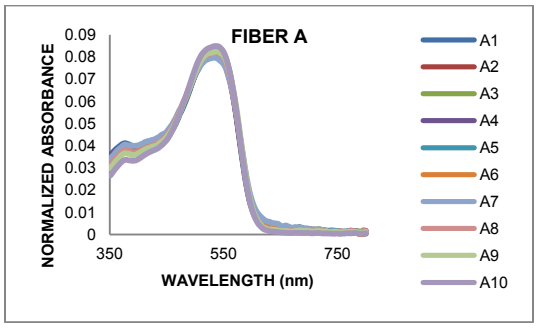
Averaged Spectra for EV Fiber E



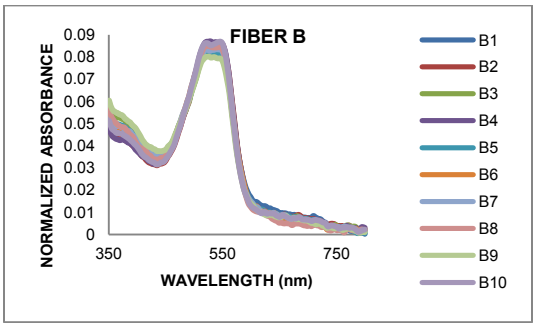
Averaged Spectra for EV Fiber F

A.2. All Dyes Analyzed at IUPUI

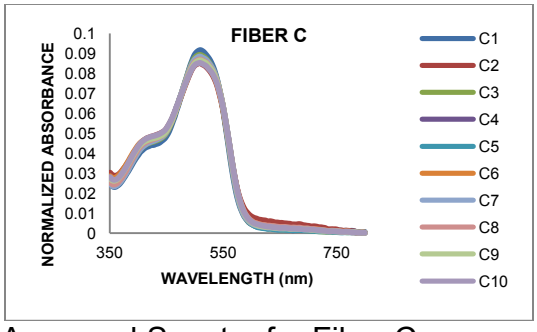
A.2.1. Training Set



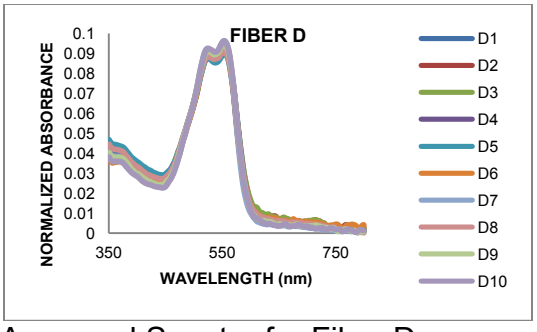
Averaged Spectra for Fiber A



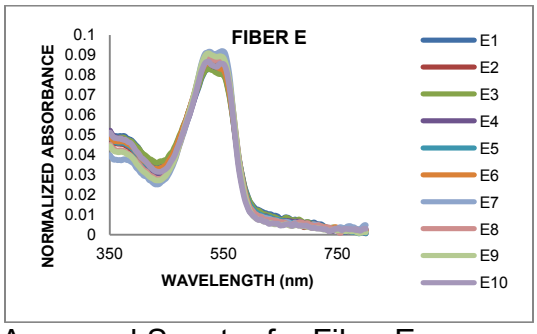
Averaged Spectra for Fiber B



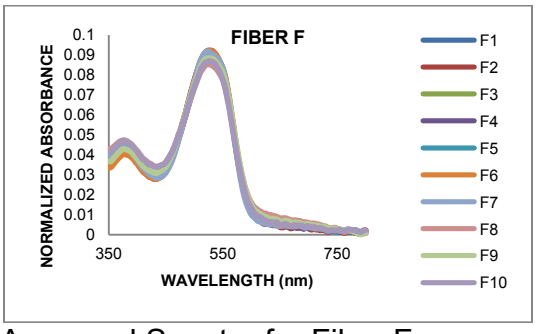
Averaged Spectra for Fiber C



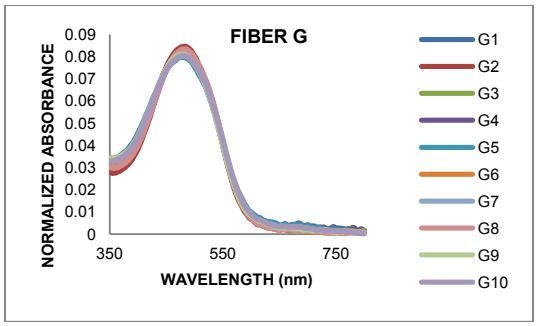
Averaged Spectra for Fiber D



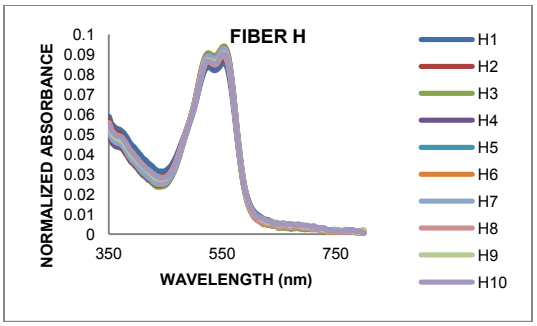
Averaged Spectra for Fiber E



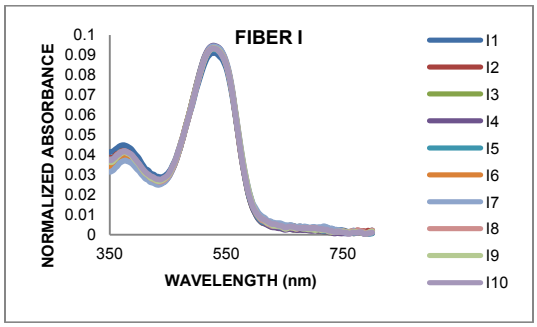
Averaged Spectra for Fiber F



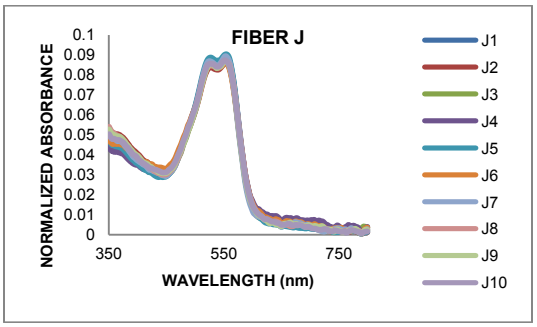
Averaged Spectra for Fiber G



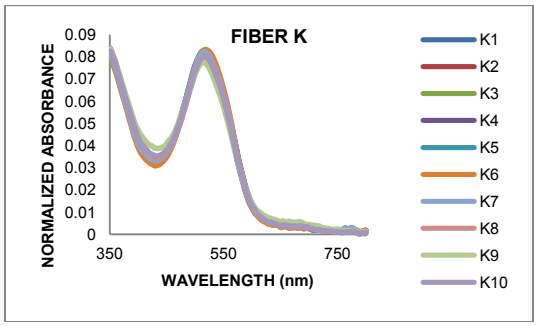
Averaged Spectra for Fiber H



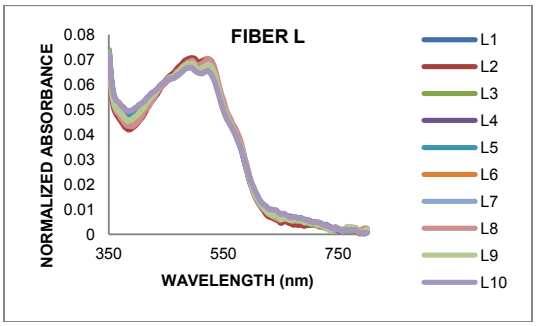
Averaged Spectra for Fiber I



Averaged Spectra for Fiber J

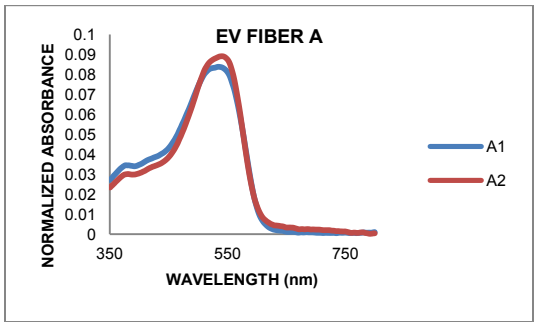


Averaged Spectra for Fiber K

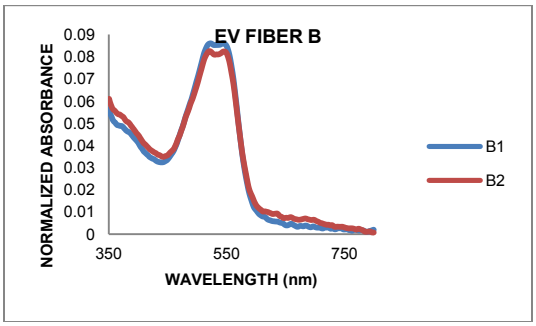


Averaged Spectra for Fiber L

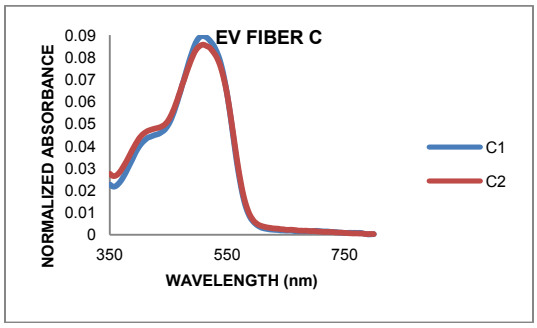
A.2.2. External Validation



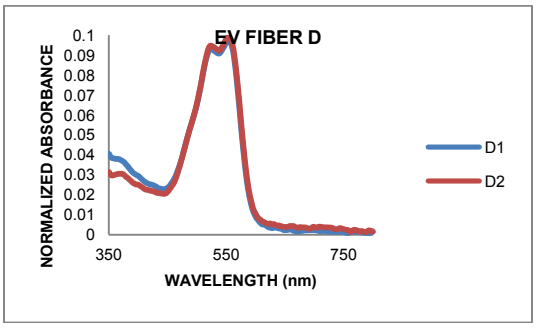
Averaged Spectra for EV Fiber A



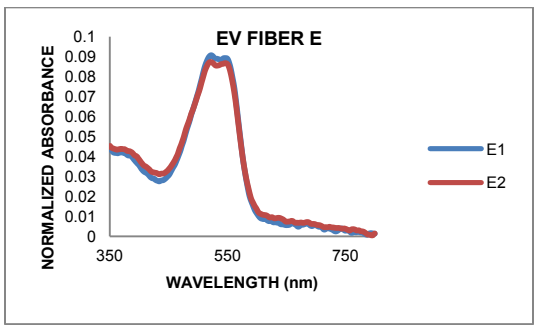
Averaged Spectra for EV Fiber B



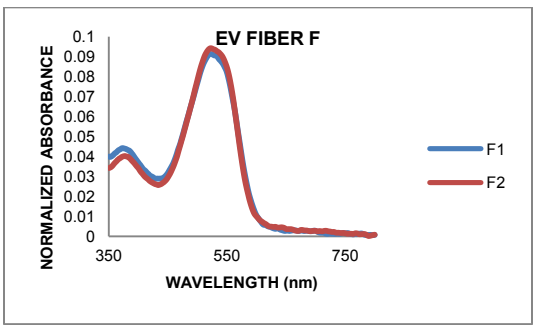
Averaged Spectra for EV Fiber C



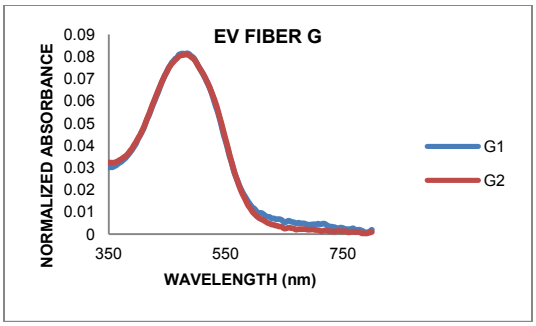
Averaged Spectra for EV Fiber D



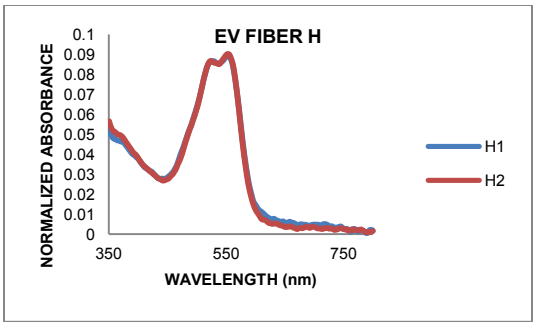
Averaged Spectra for EV Fiber E



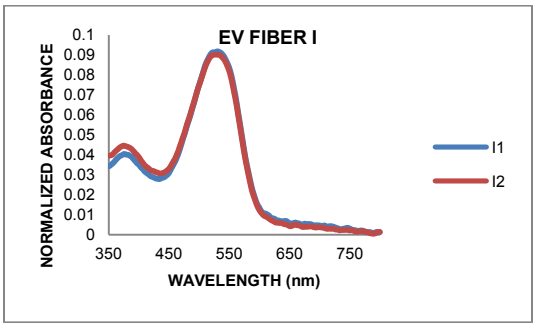
Averaged Spectra for EV Fiber F



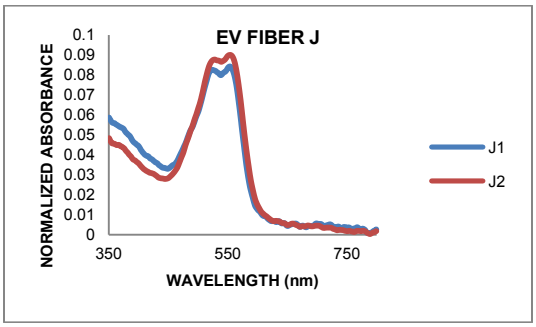
Averaged Spectra for EV Fiber G



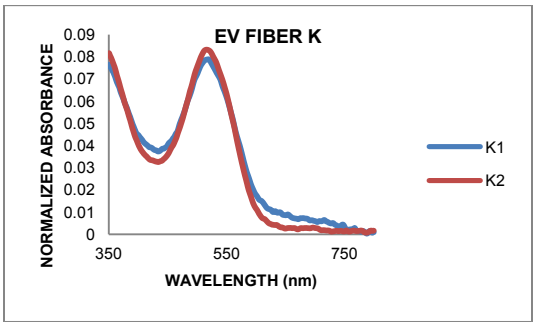
Averaged Spectra for EV Fiber H



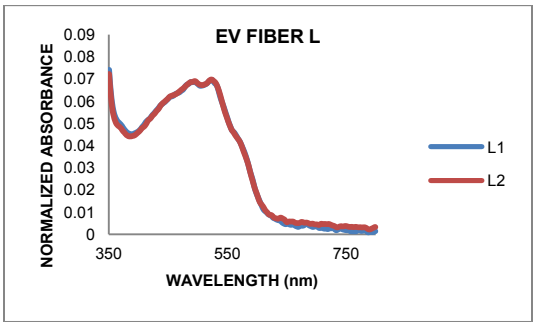
Averaged Spectra for EV Fiber I



Averaged Spectra for EV Fiber J

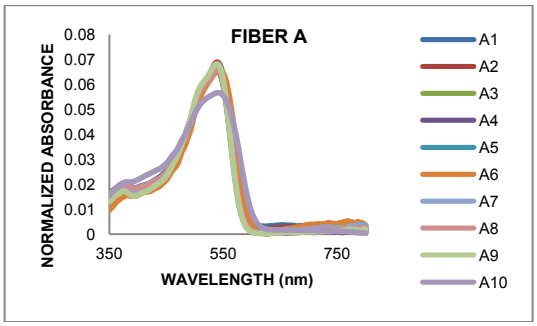


Averaged Spectra for EV Fiber K

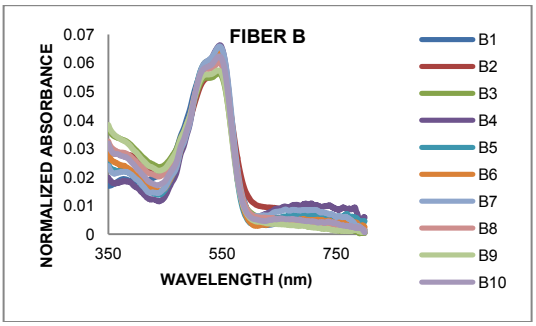


Averaged Spectra for EV Fiber L

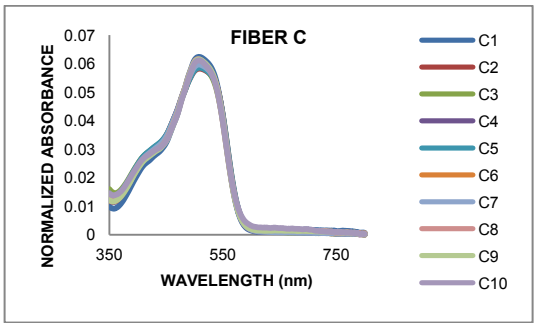
A.3. All Dyes Analyzed at ISP



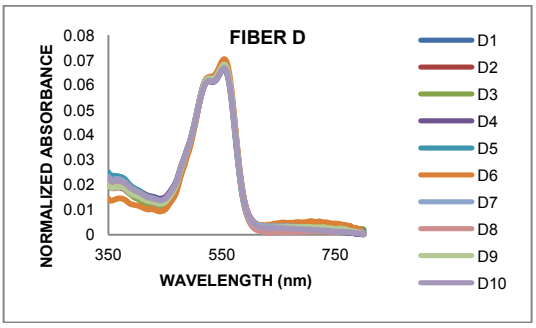
Averaged Spectra for Fiber A



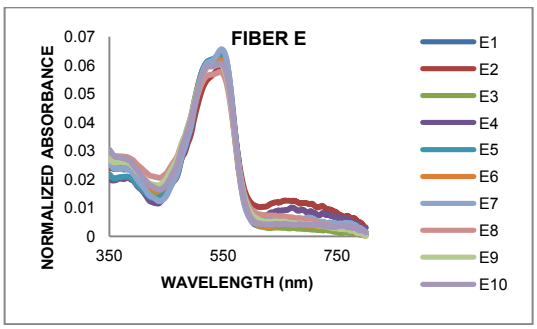
Averaged Spectra for Fiber B



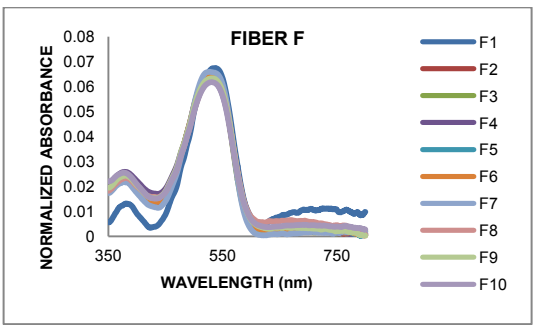
Averaged Spectra for Fiber C



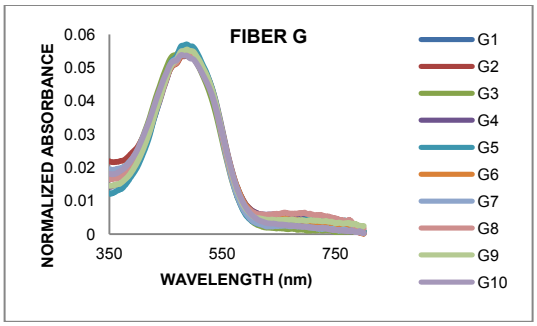
Averaged Spectra for Fiber D



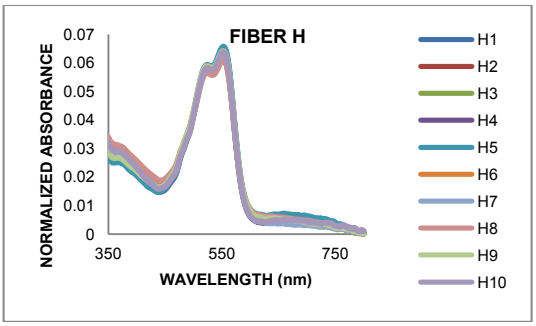
Averaged Spectra for Fiber E



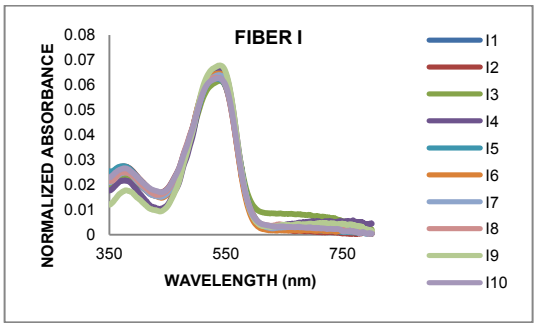
Averaged Spectra for Fiber F



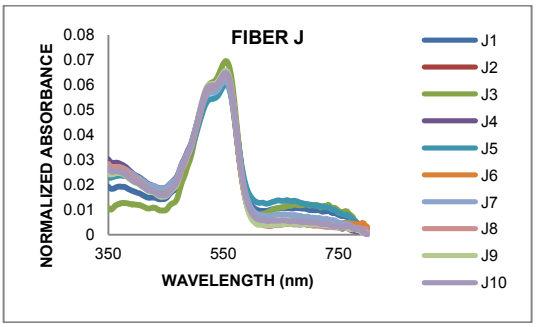
Averaged Spectra for Fiber G



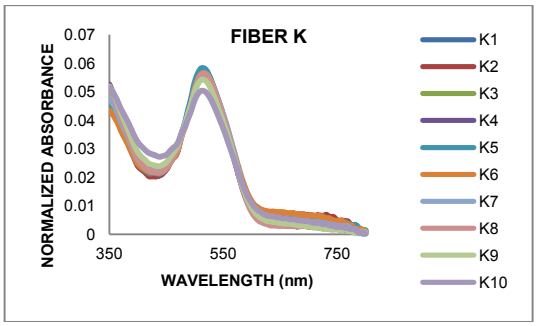
Averaged Spectra for Fiber H



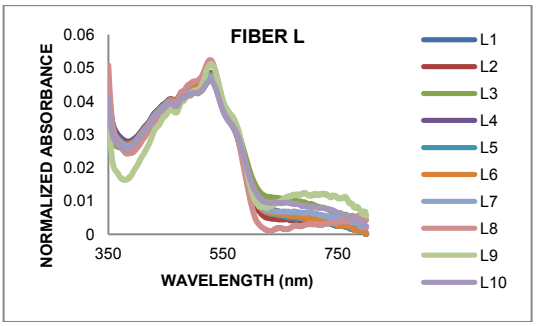
Averaged Spectra for Fiber I



Averaged Spectra for Fiber J



Averaged Spectra for Fiber K

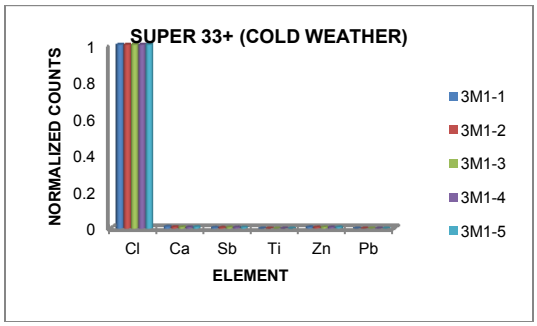


Averaged Spectra for Fiber L

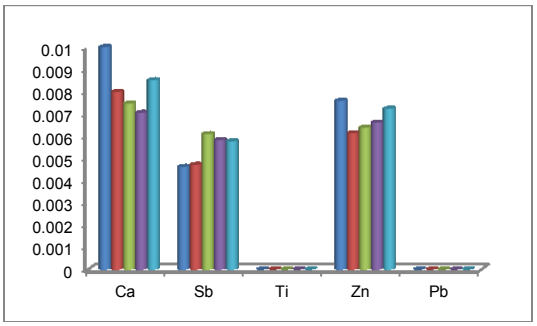
Appendix B. Tape Backings

B.1. XRF Analysis

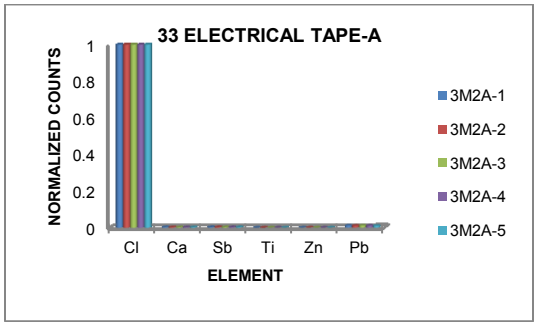
B.1.1. Black Adhesives



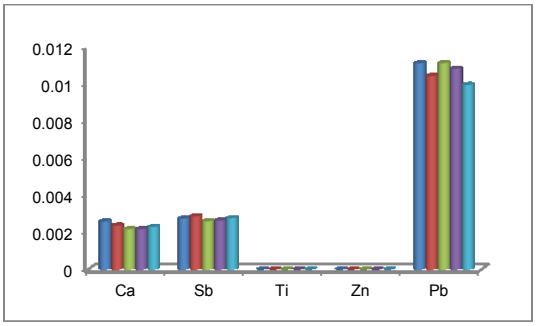
Elemental Composition of 3M1



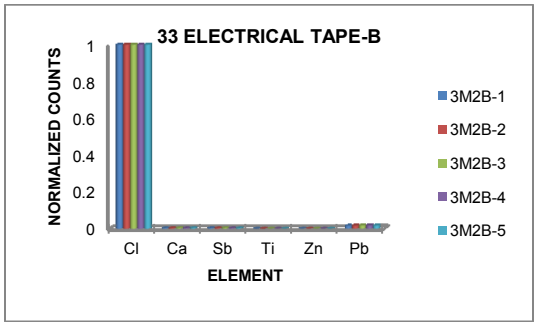
Composition Without Cl for 3M1



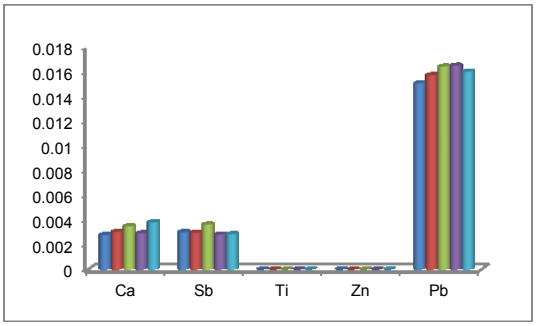
Elemental Composition of 3M2A



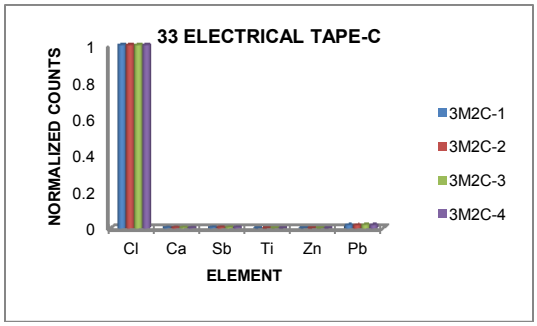
Composition Without Cl for 3M2A



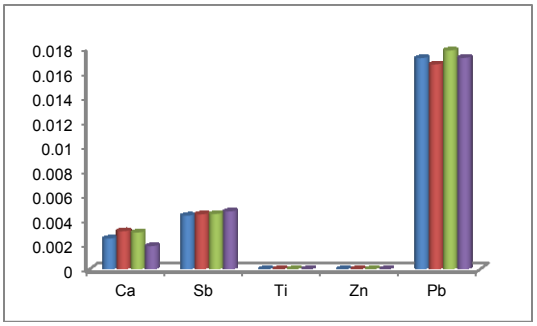
Elemental Composition of 3M2B



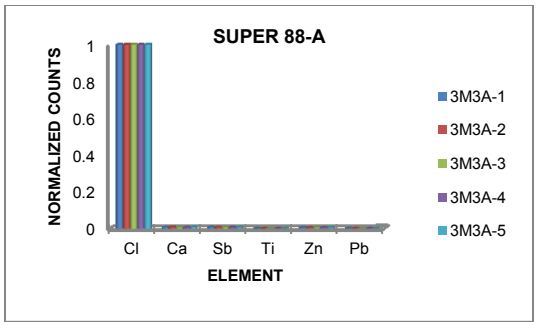
Composition Without Cl for 3M2B



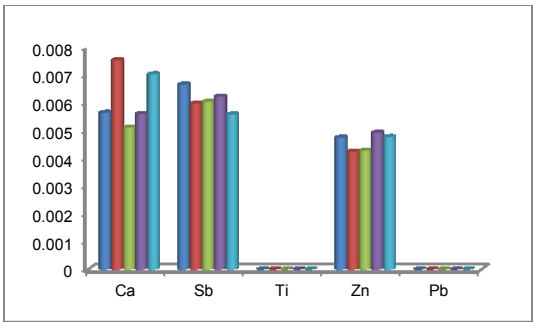
Elemental Composition of 3M2C



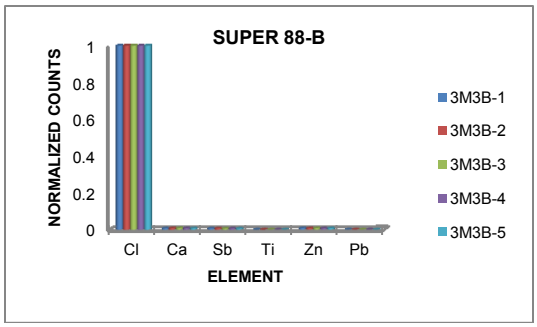
Composition Without Cl for 3M2C



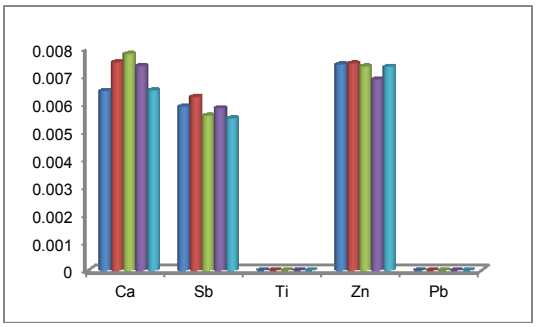
Elemental Composition of 3M3A



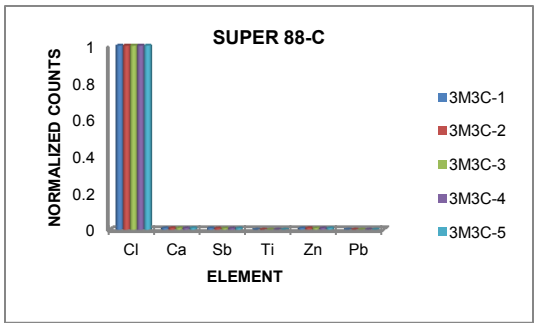
Composition Without Cl for 3M3A



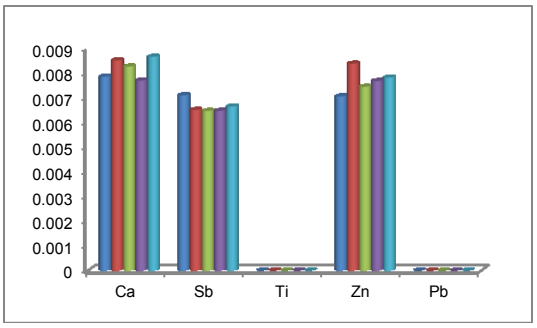
Elemental Composition of 3M3B



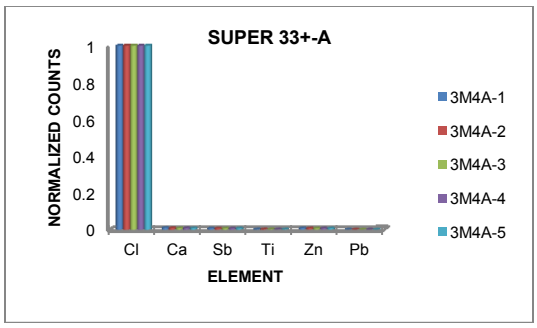
Composition Without Cl for 3M3B



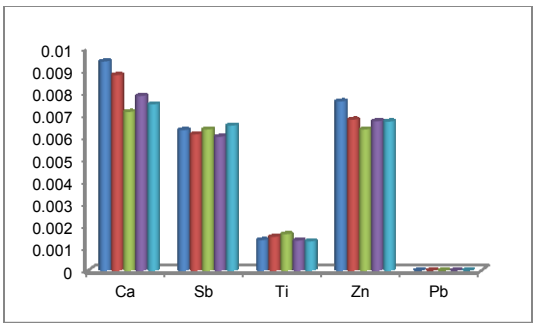
Elemental Composition of 3M3C



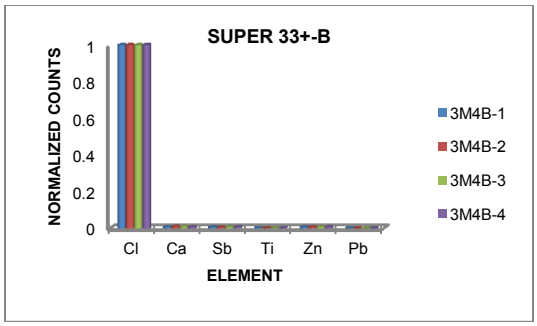
Composition Without Cl for 3M3C



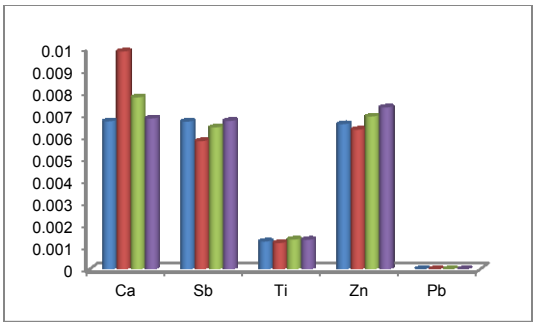
Elemental Composition of 3M4A



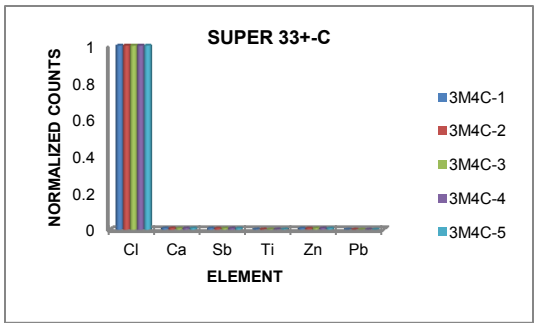
Composition Without Cl for 3M4A



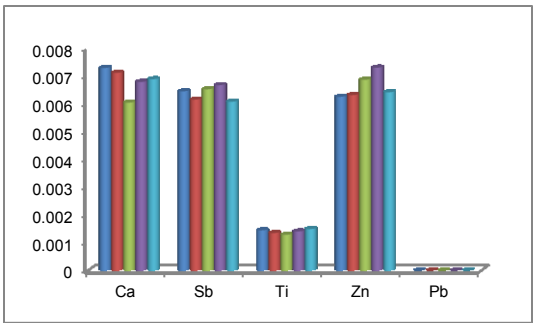
Elemental Composition of 3M4B



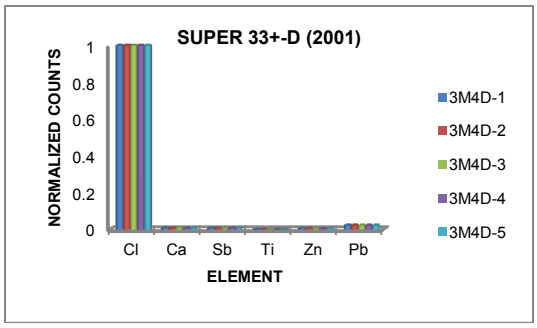
Composition Without Cl for 3M4B



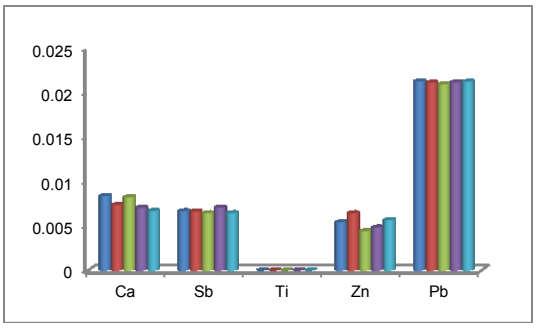
Elemental Composition of 3M4C



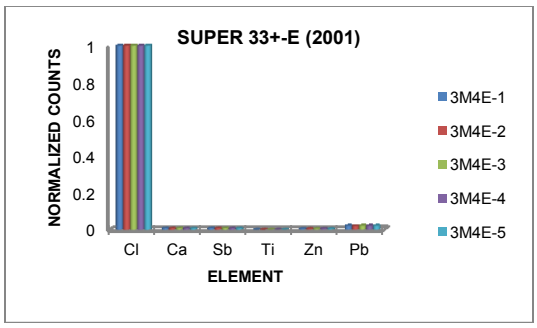
Composition Without Cl for 3M4C



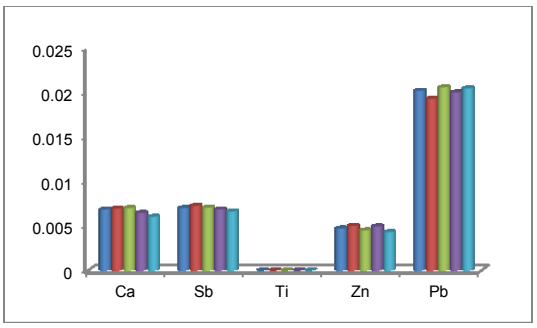
Elemental Composition of 3M4D



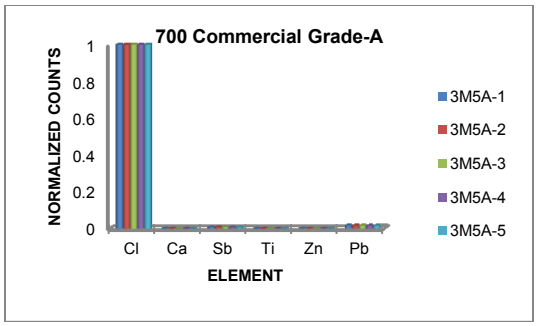
Composition Without Cl for 3M4D



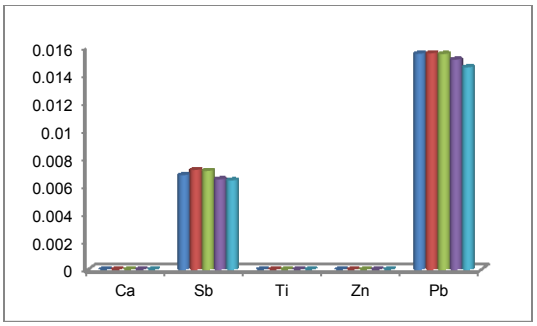
Elemental Composition of 3M4E



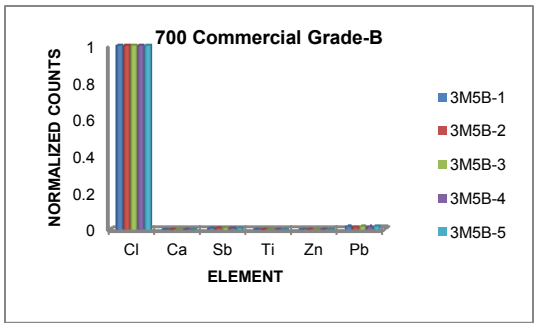
Composition Without Cl for 3M4E



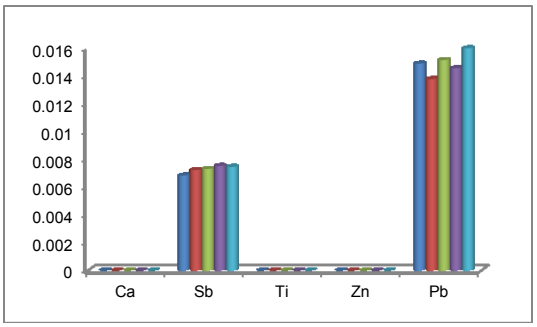
Elemental Composition of 3M5A



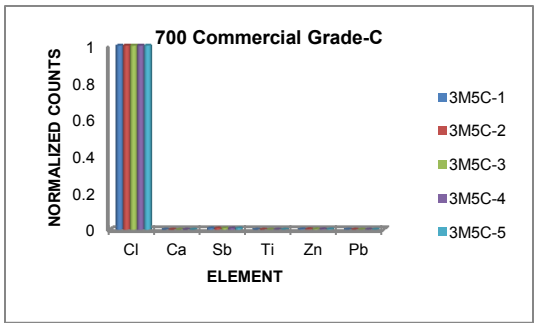
Composition Without Cl for 3M5A



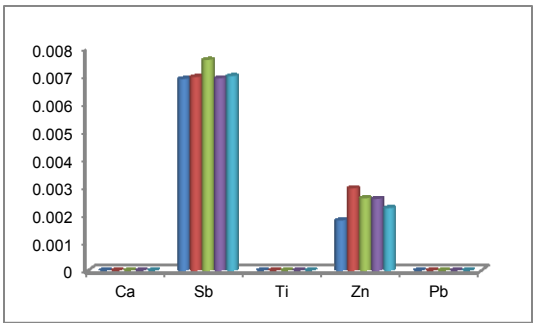
Elemental Composition of 3M5B



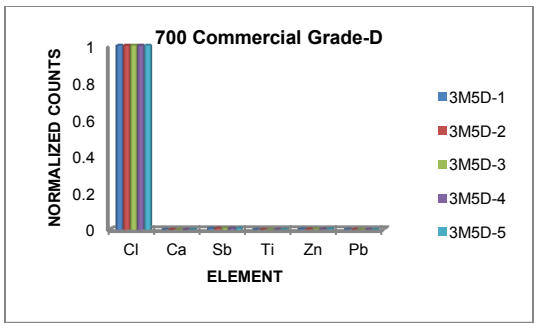
Composition Without Cl for 3M5B



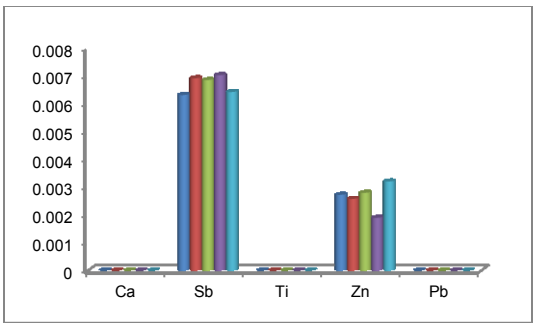
Elemental Composition of 3M5C



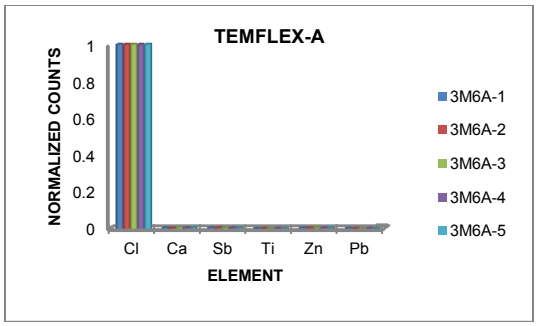
Composition Without Cl for 3M5C



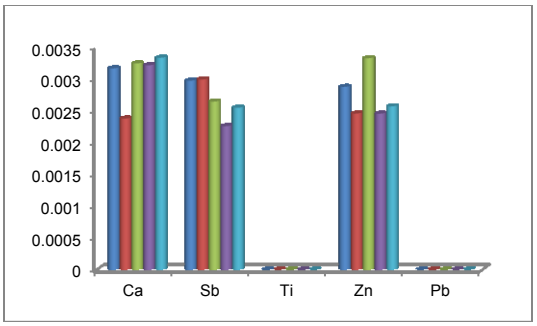
Elemental Composition of 3M5D



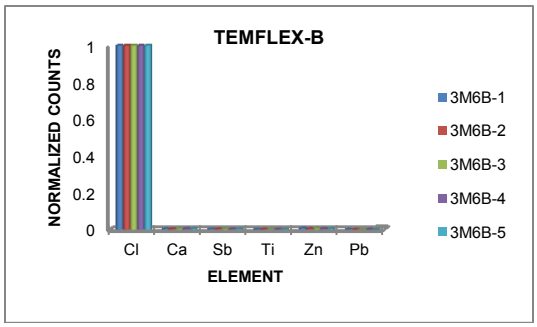
Composition Without Cl for 3M5D



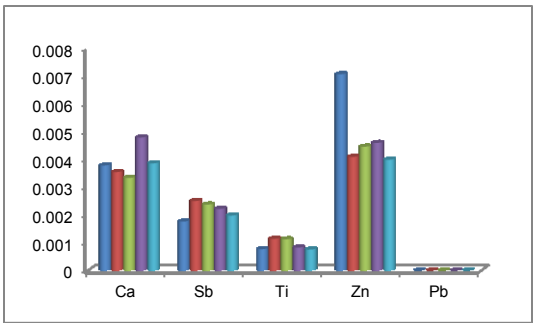
Elemental Composition of 3M6A



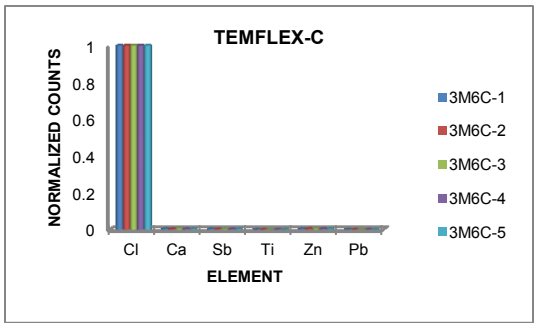
Composition Without Cl for 3M6A



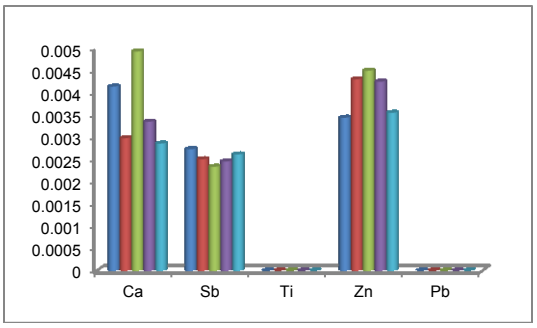
Elemental Composition of 3M6B



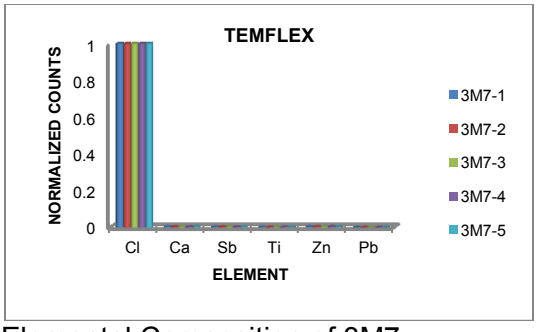
Composition Without Cl for 3M6B



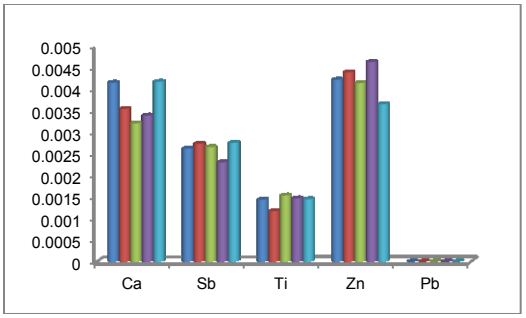
Elemental Composition of 3M6C



Composition Without Cl for 3M6C

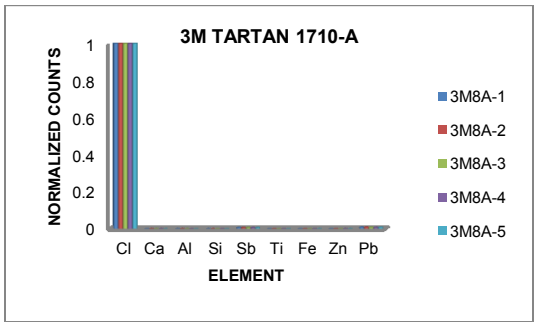


Elemental Composition of 3M7

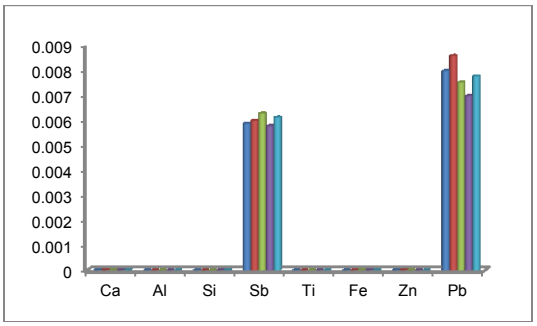


Composition Without Cl for 3M7

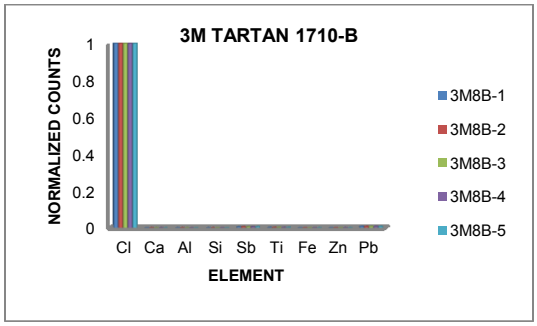
B.1.2. Colorless Adhesives



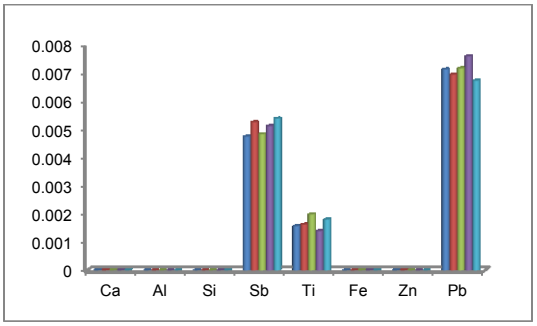
Elemental Composition of 3M8A



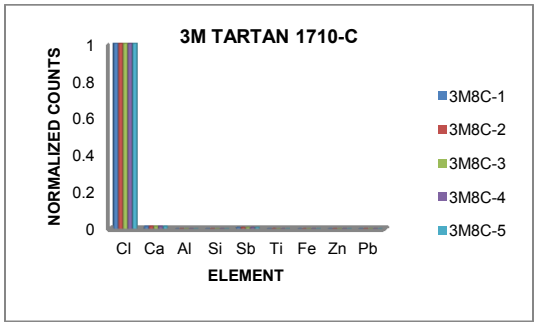
Composition Without Cl for 3M8A



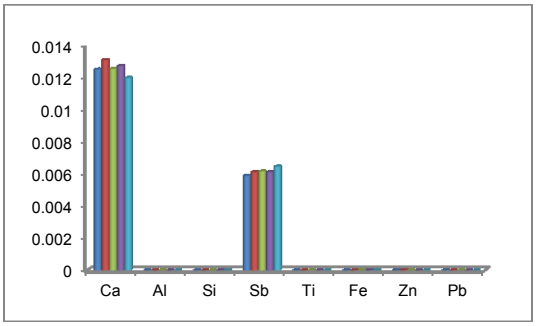
Elemental Composition of 3M8B



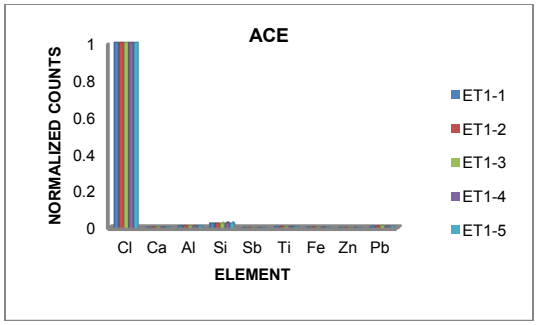
Composition Without Cl for 3M8B



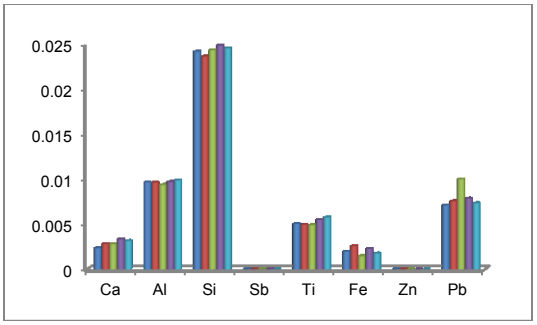
Elemental Composition of 3M8C



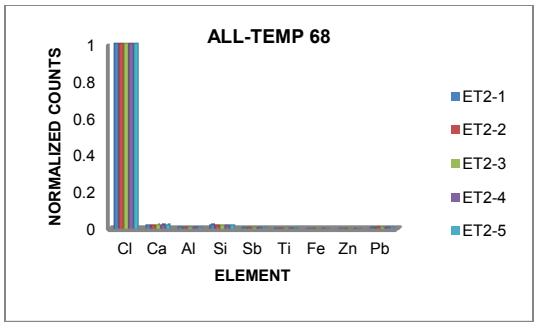
Composition Without Cl for 3M8C



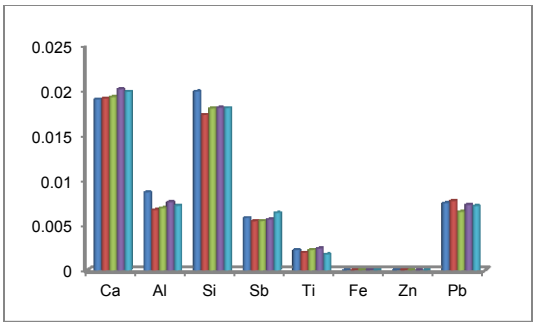
Elemental Composition of ET1



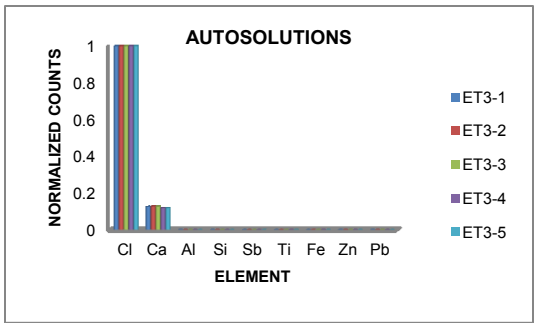
Composition Without Cl for ET1



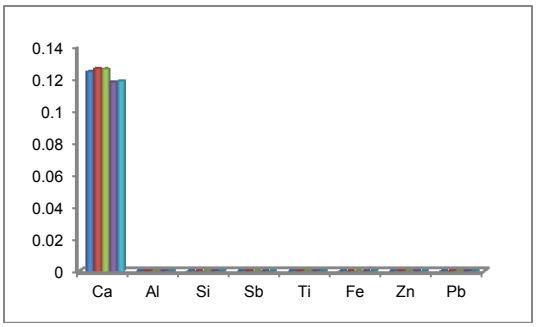
Elemental Composition of ET2



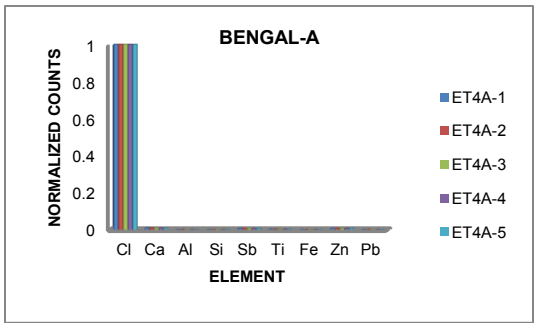
Composition Without Cl for ET2



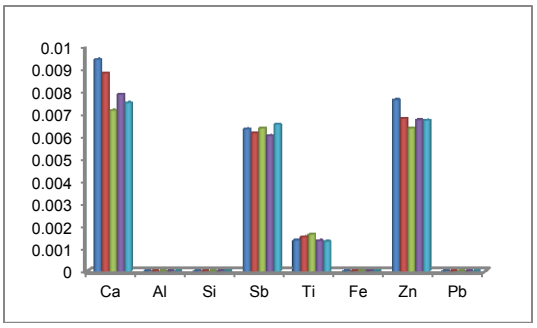
Elemental Composition of ET3



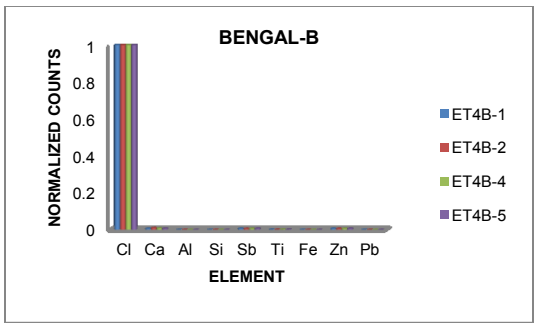
Composition Without Cl for ET3



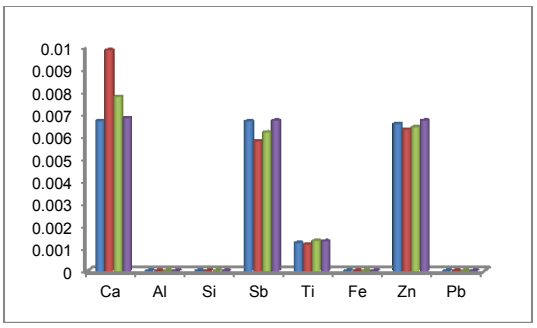
Elemental Composition of ET4A



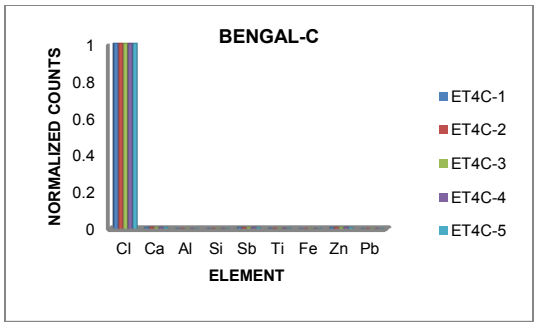
Composition Without Cl for ET4A



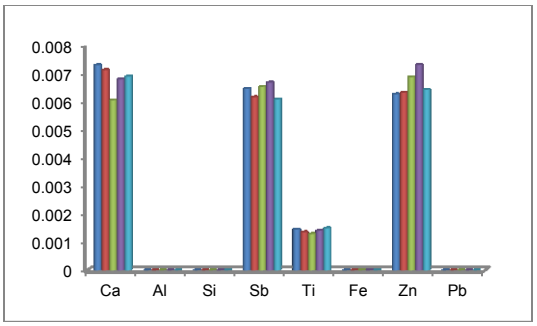
Elemental Composition of ET4B



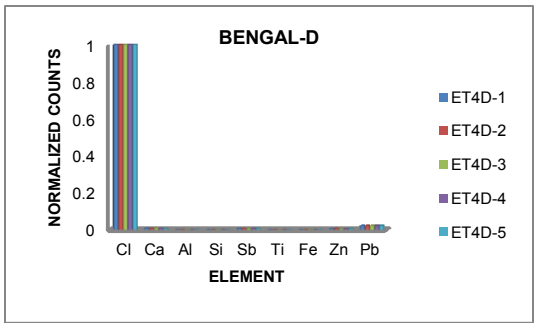
Composition Without Cl for ET4B



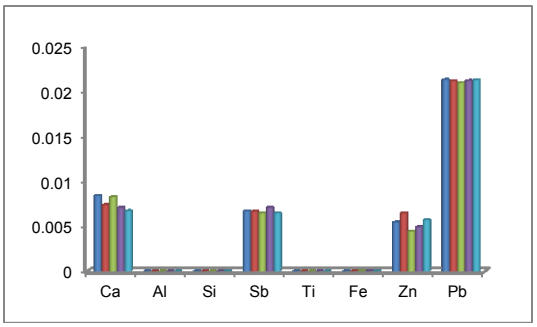
Elemental Composition of ET4C



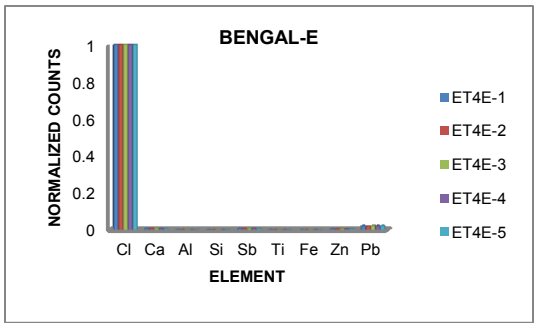
Composition Without Cl for ET4C



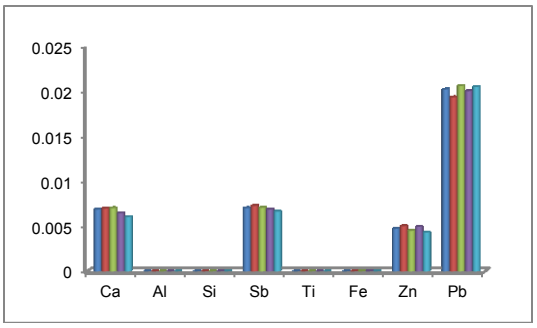
Elemental Composition of ET4D



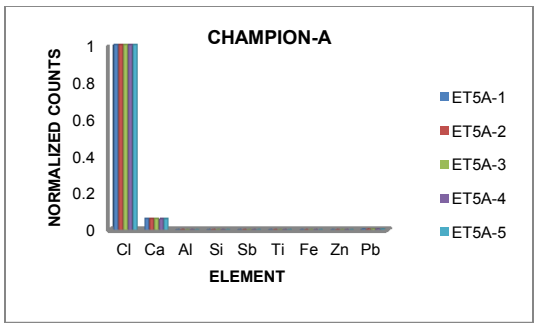
Composition Without Cl for ET4D



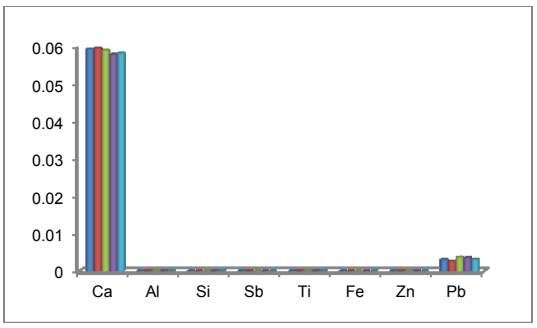
Elemental Composition of ET4E



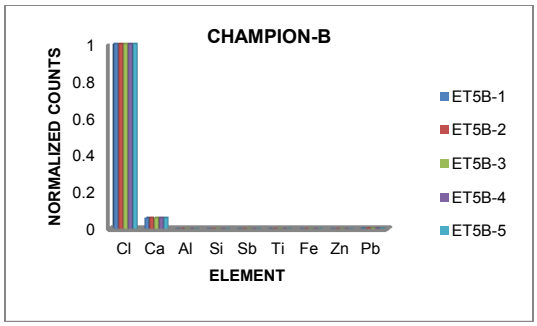
Composition Without Cl for ET4E



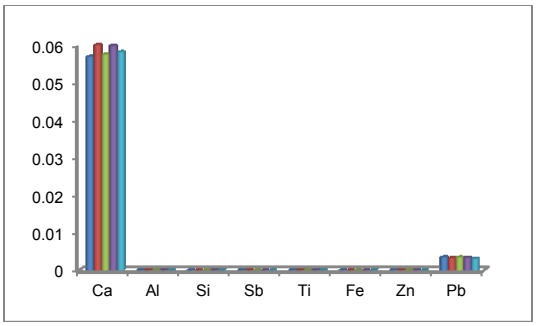
Elemental Composition of ET5A



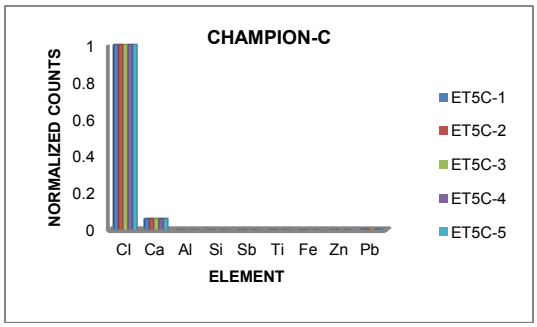
Composition Without Cl for ET5A



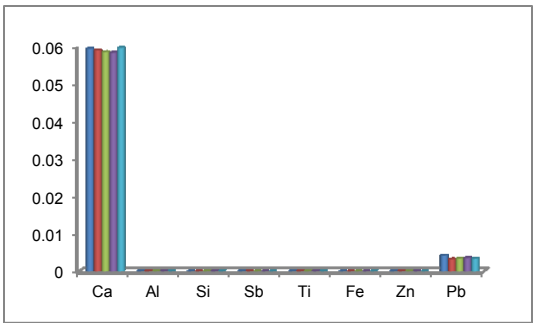
Elemental Composition of ET5B



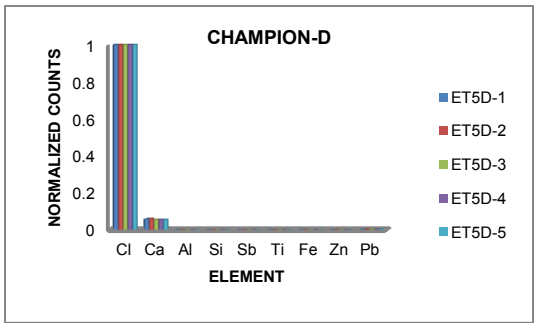
Composition Without Cl for ET5B



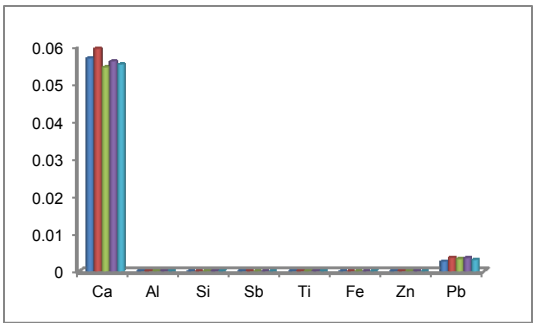
Elemental Composition of ET5C



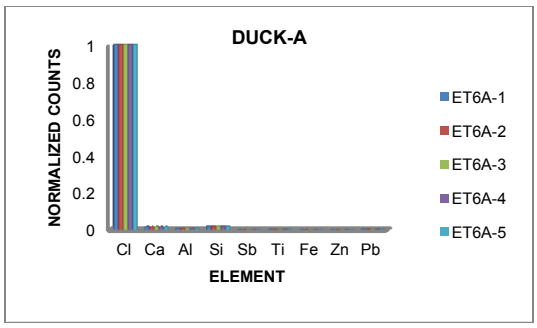
Composition Without Cl for ET5C



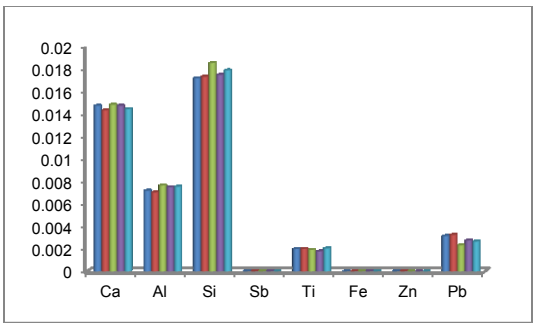
Elemental Composition of ET5D



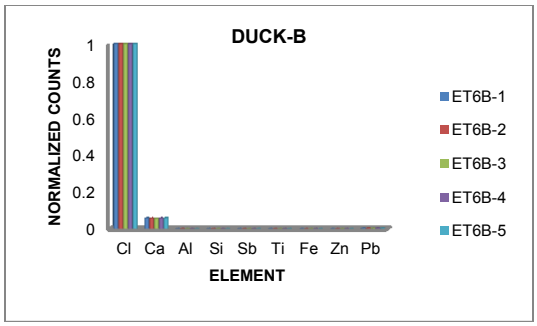
Composition Without Cl for ET5D



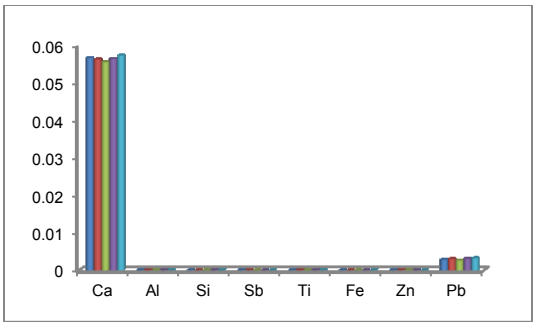
Elemental Composition of ET6A



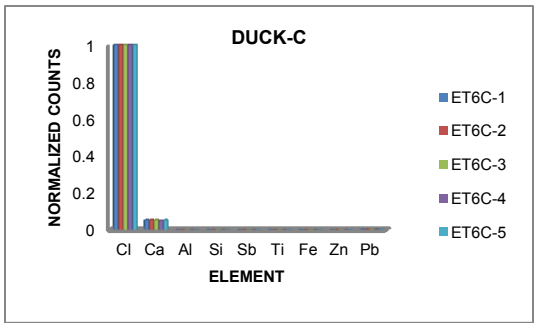
Composition Without Cl for ET6A



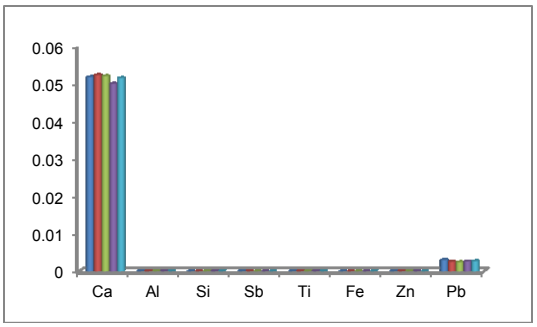
Elemental Composition of ET6B



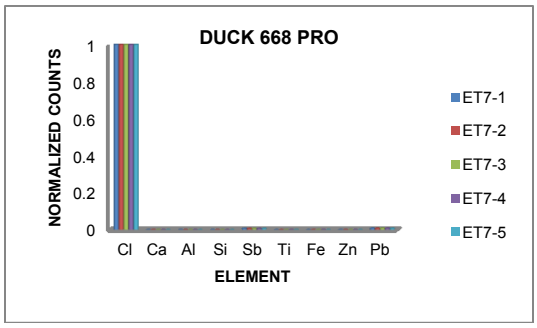
Composition Without Cl for ET6B



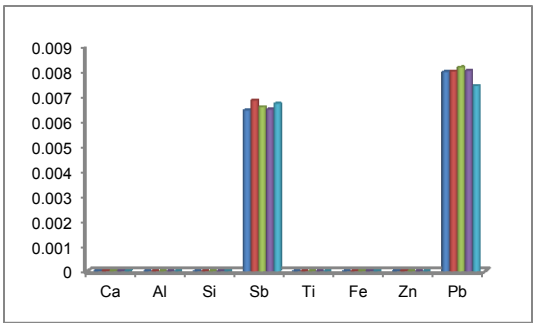
Elemental Composition of ET6C



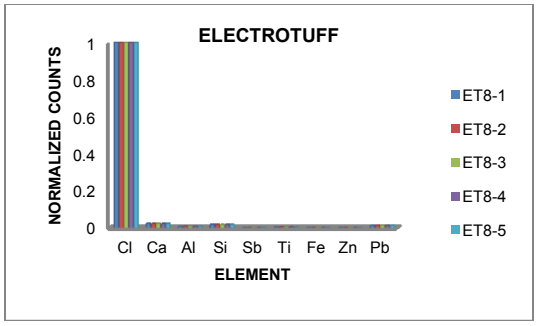
Composition Without Cl for ET6C



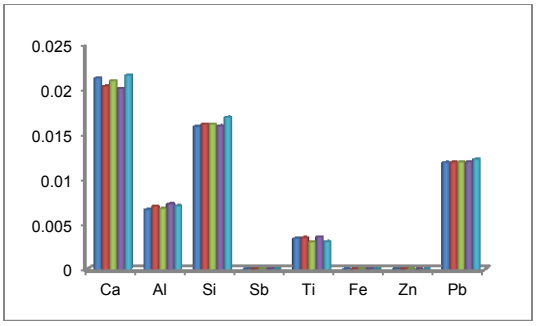
Elemental Composition of ET7



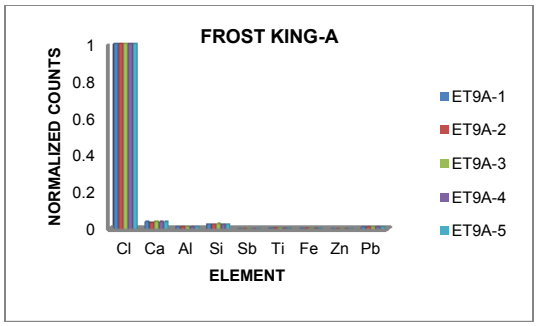
Composition Without Cl for ET7



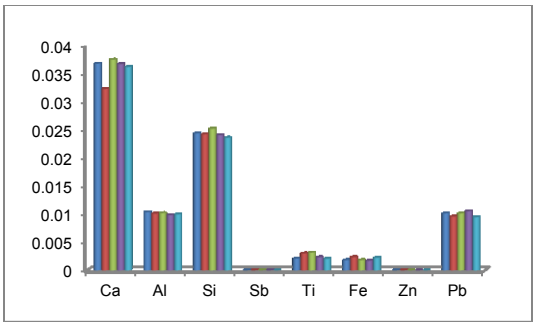
Elemental Composition of ET8



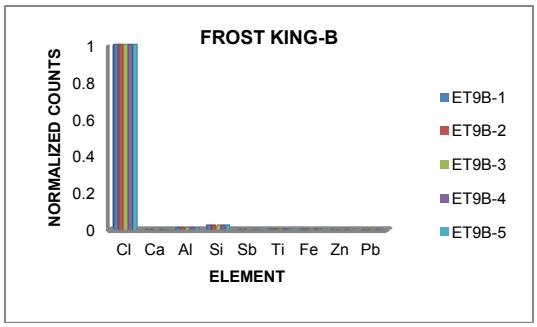
Composition Without Cl for ET8



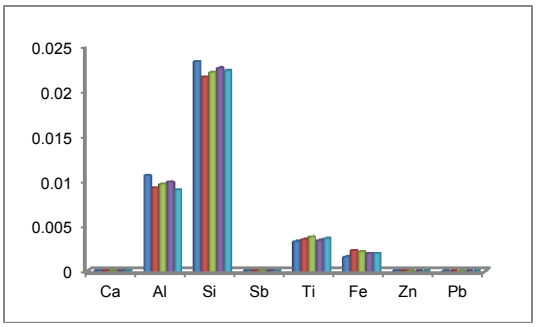
Elemental Composition of ET9A



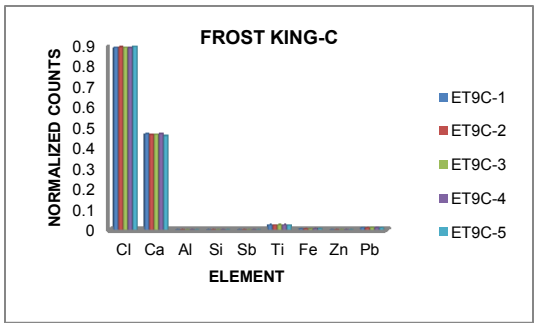
Composition Without Cl for ET9A



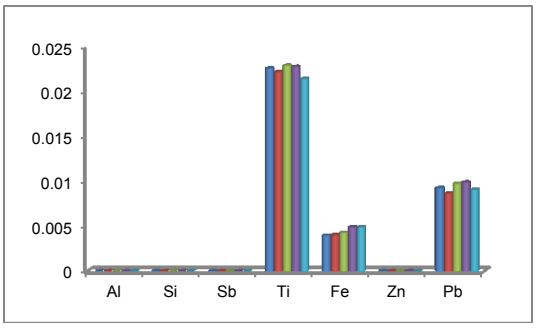
Elemental Composition of ET9B



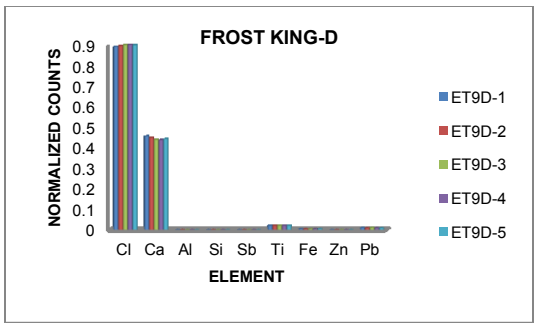
Composition Without Cl for ET9B



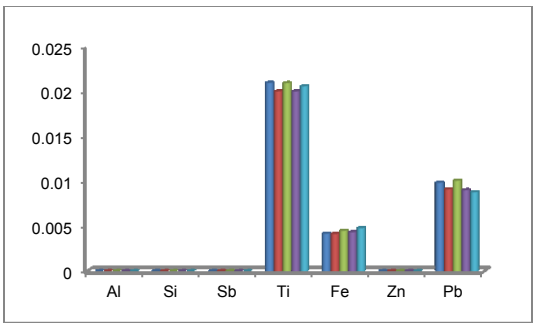
Elemental Composition of ET9C



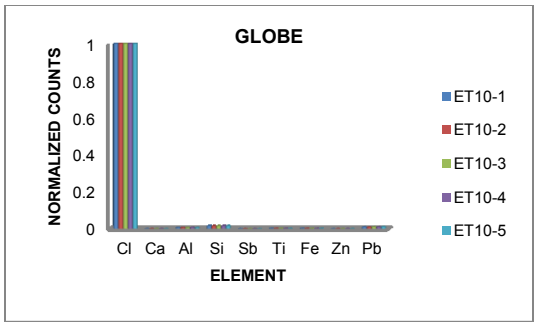
Composition Without Cl & Ca for ET9C



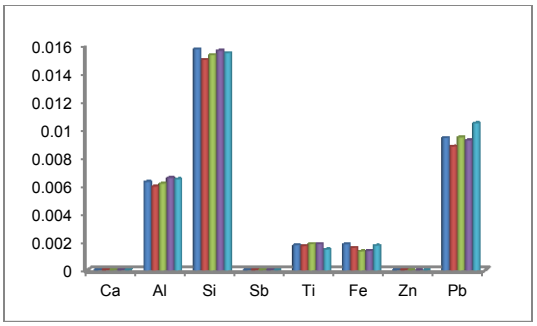
Elemental Composition of ET9D



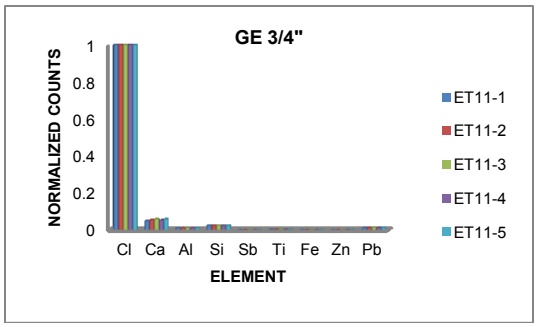
Composition Without Cl & Ca for ET9D



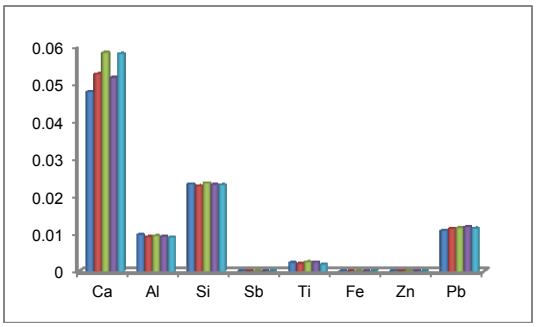
Elemental Composition of ET10



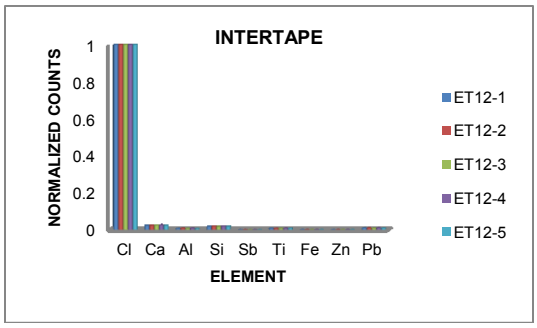
Composition Without Cl for ET10



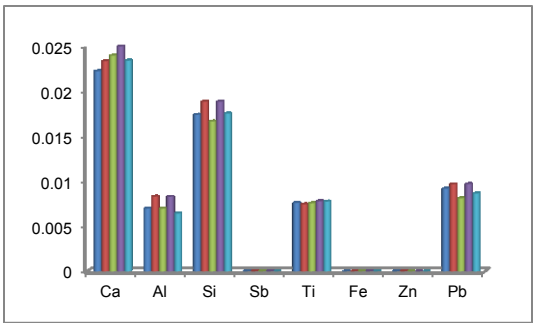
Elemental Composition of ET11



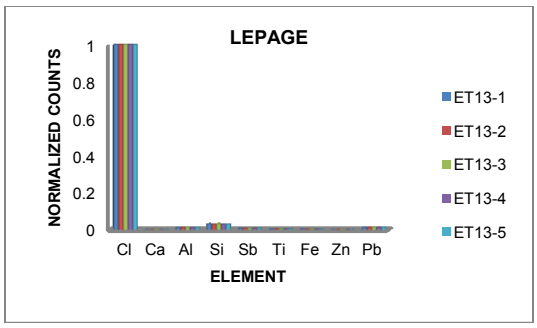
Composition Without Cl for ET11



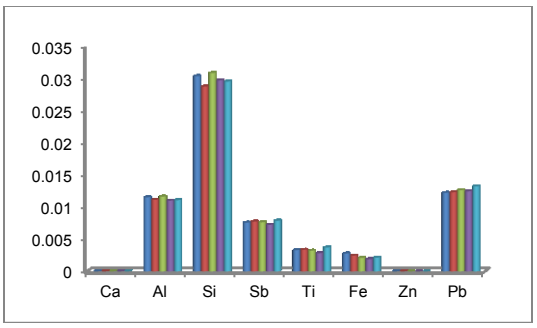
Elemental Composition of ET12



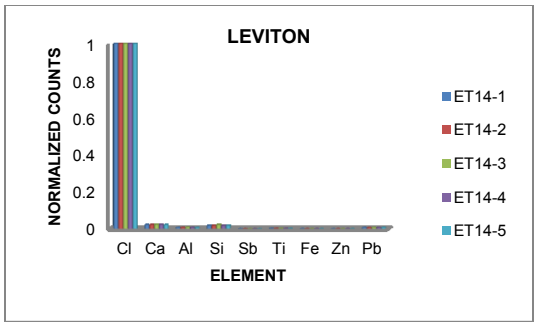
Composition Without Cl for ET12



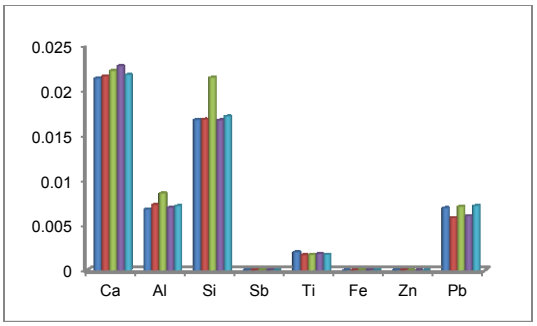
Elemental Composition of ET13



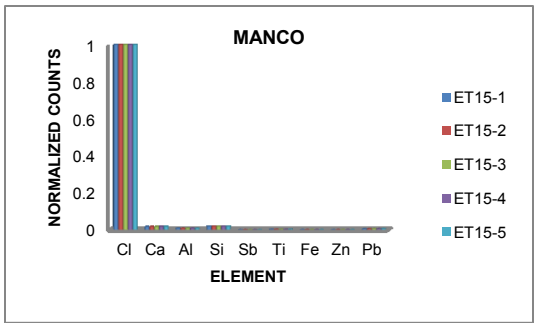
Composition Without Cl for ET13



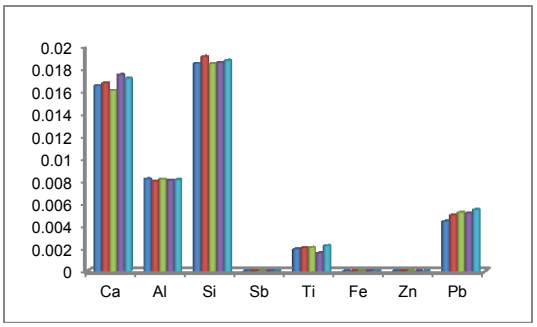
Elemental Composition of ET14



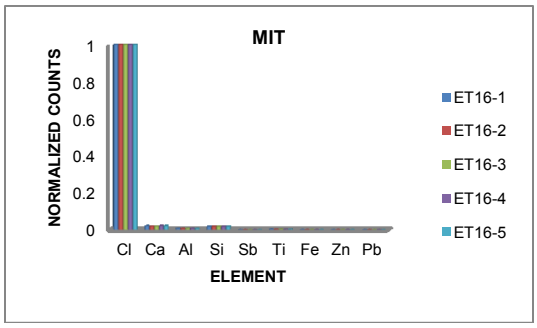
Composition Without Cl for ET14



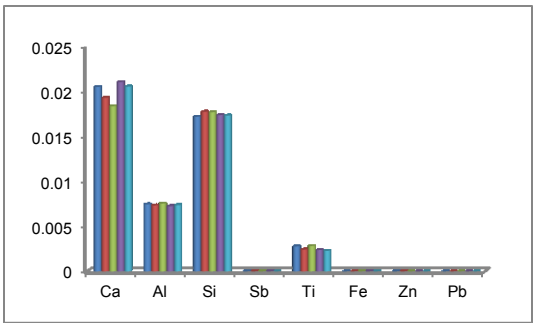
Elemental Composition of ET15



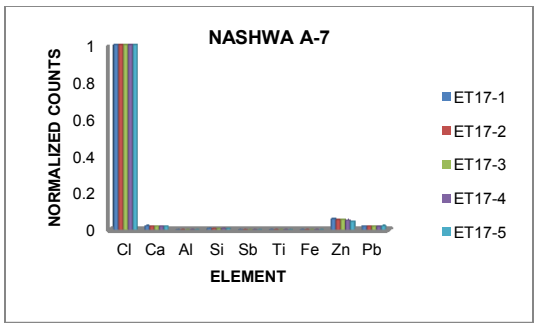
Composition Without Cl for ET15



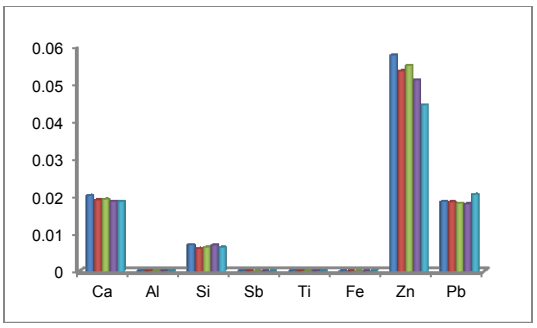
Elemental Composition of ET16



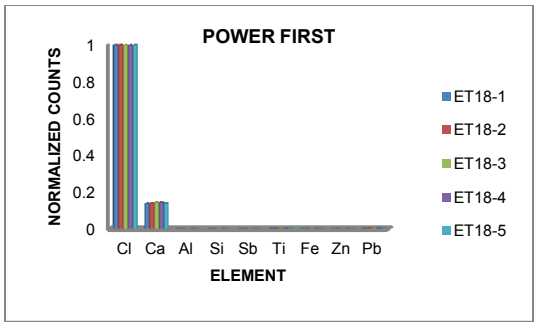
Composition Without Cl for ET16



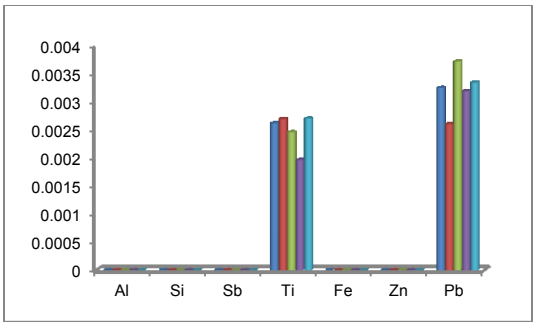
Elemental Composition of ET17



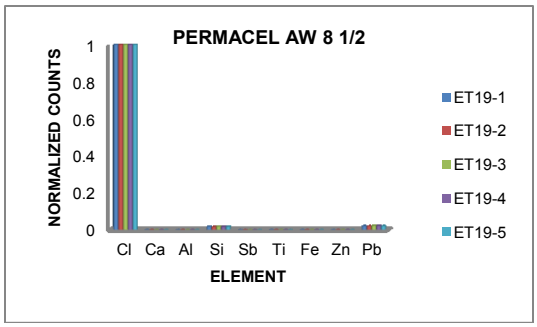
Composition Without Cl for ET17



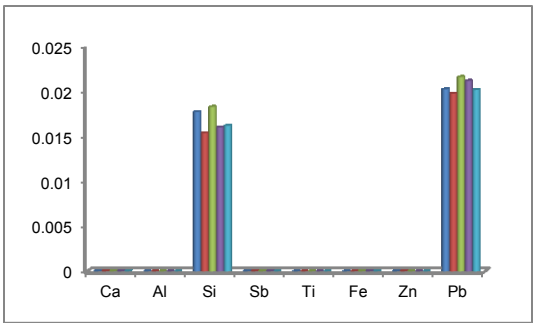
Elemental Composition of ET18



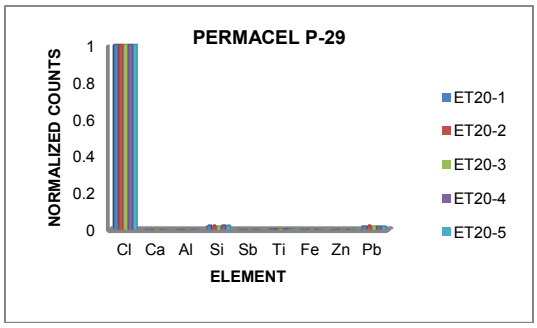
Composition Without Cl for ET18



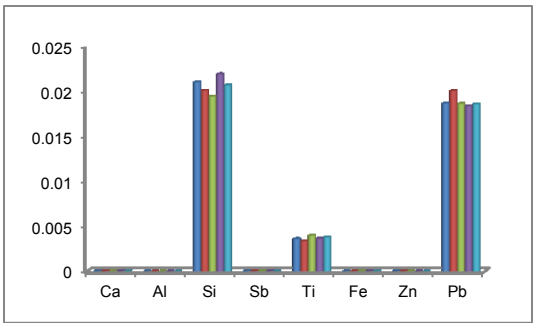
Elemental Composition of ET19



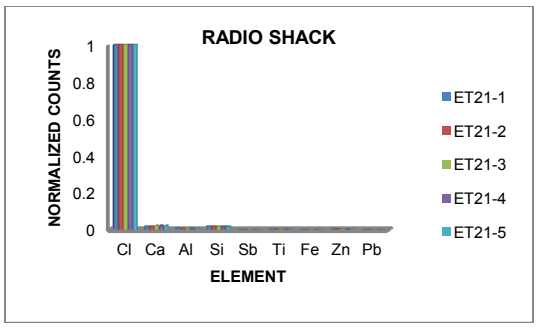
Composition Without Cl for ET19



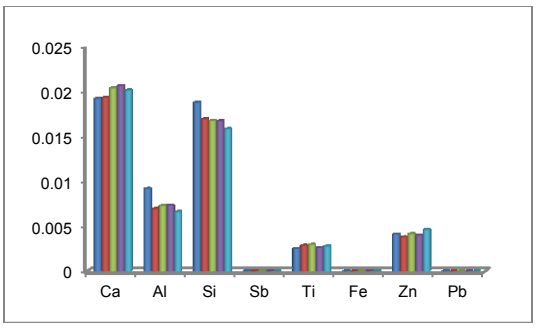
Elemental Composition of ET20



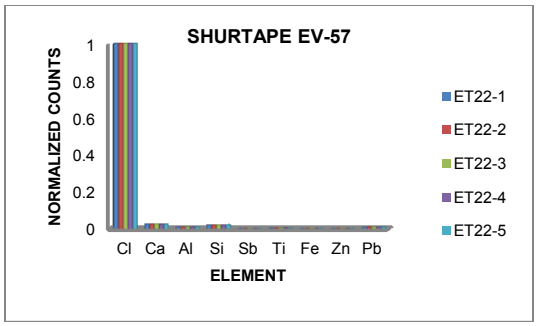
Composition Without Cl for ET20



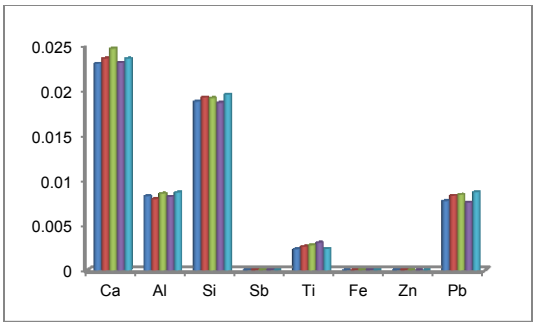
Elemental Composition of ET21



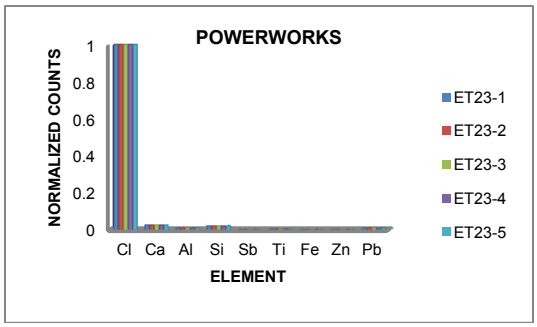
Composition Without Cl for ET21



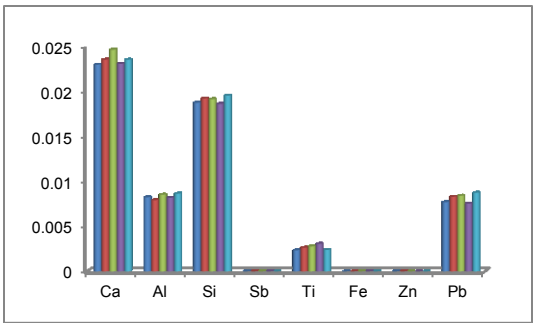
Elemental Composition of ET22



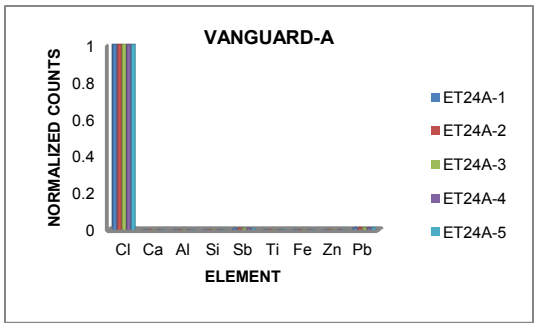
Composition Without Cl for ET22



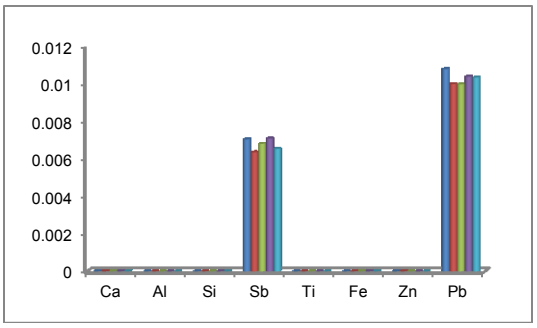
Elemental Composition of ET23



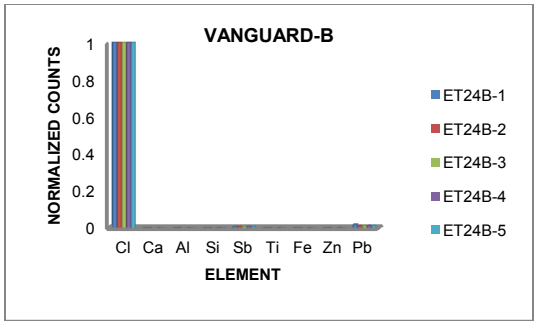
Composition Without Cl for ET23



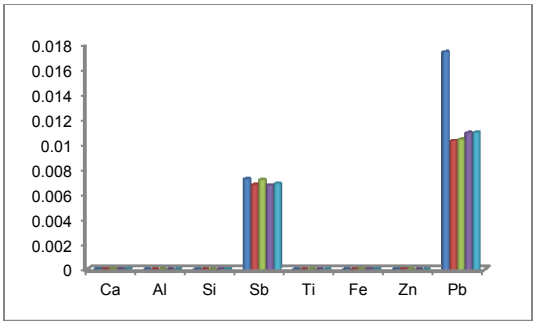
Elemental Composition of ET24A



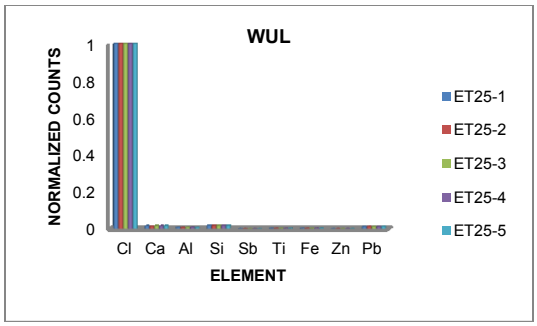
Composition Without Cl for ET24A



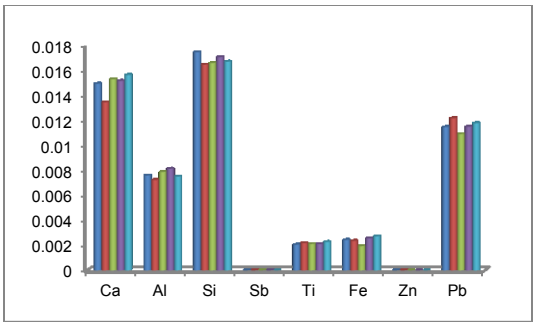
Elemental Composition of ET24B



Composition Without Cl for ET24B



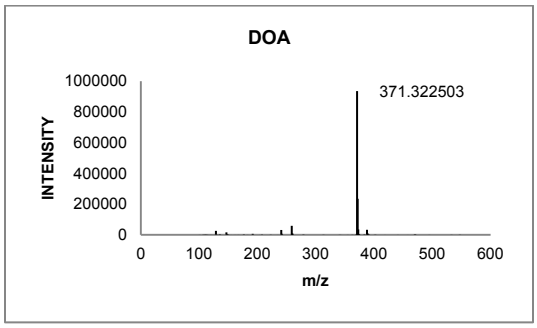
Elemental Composition of ET25



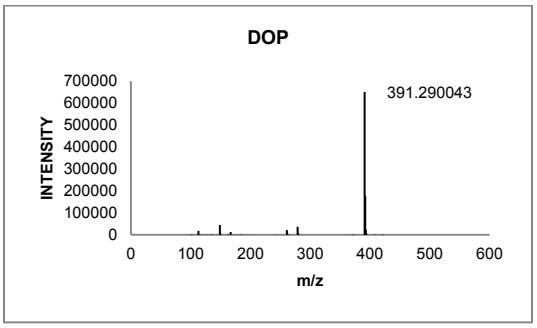
Composition Without Cl for ET25

B.2. DART/MS Positive Ion Mode Analysis

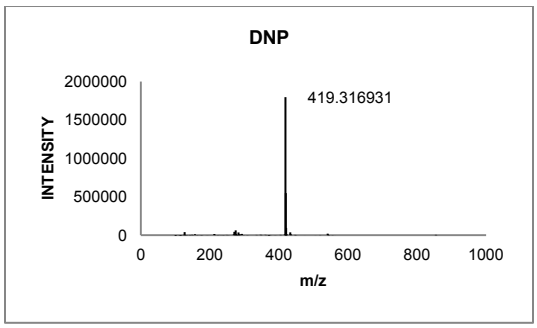
B.2.1. Plasticizer Standards



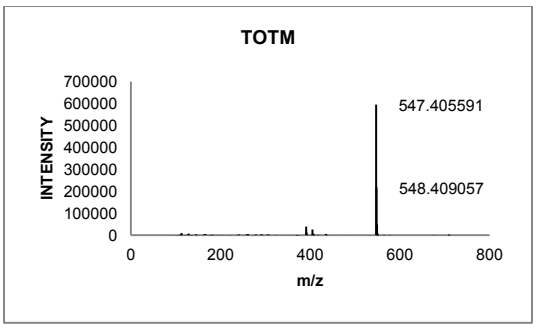
Mass Spectrum for Dioctyl adipate



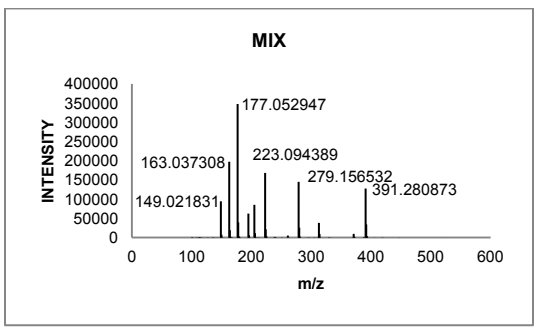
Mass Spectrum for Dioctyl phthalate



Mass Spectrum for Di-n-nonyl phthalate



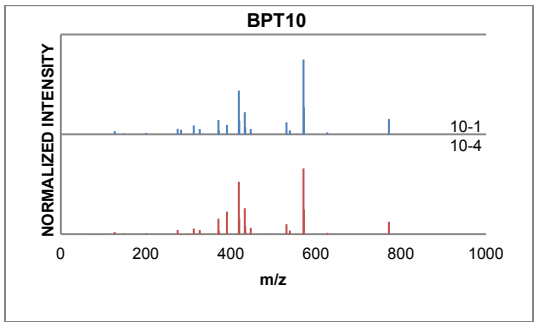
Mass Spectrum for mixture



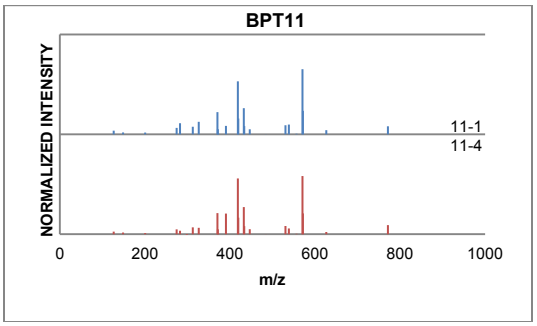
Mass Spectrum for the mixture

Mixture contains 2000µL of:
Dioctyl adipate,
Dioctyl phthalate,
Dimethyl phthalate,
Diethyl phthalate,
Butyl benzyl phthalate, and
Di-n-butyl phthalate,
in isooctane.

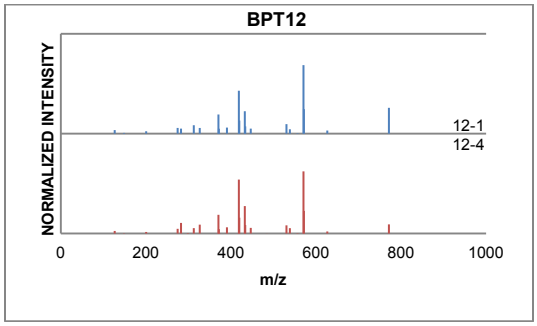
B.2.2. Black Adhesives



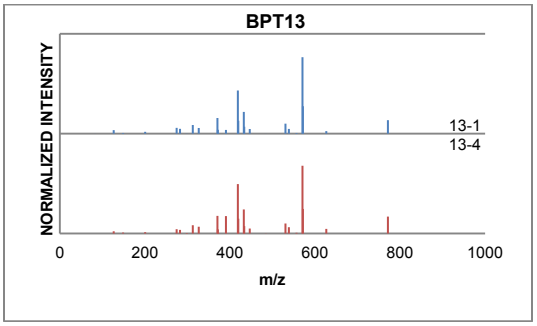
Averaged Mass Spectrum for BPT10



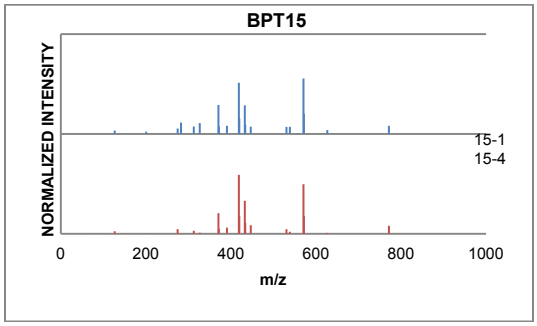
Averaged Mass Spectrum for BPT11



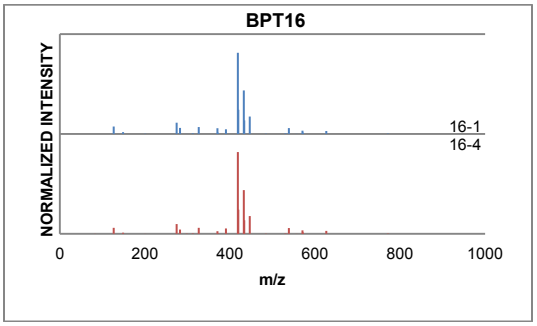
Averaged Mass Spectrum for BPT12



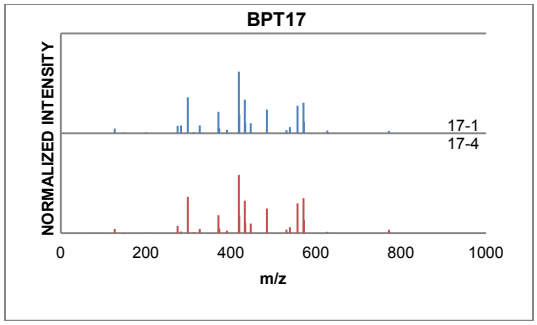
Averaged Mass Spectrum for BPT13



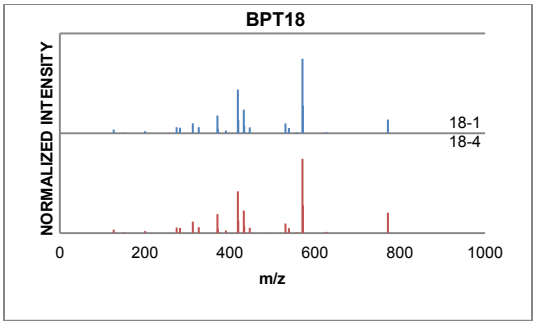
Averaged Mass Spectrum for BPT15



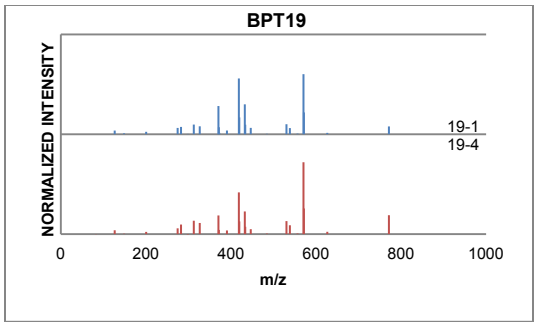
Averaged Mass Spectrum for BPT16



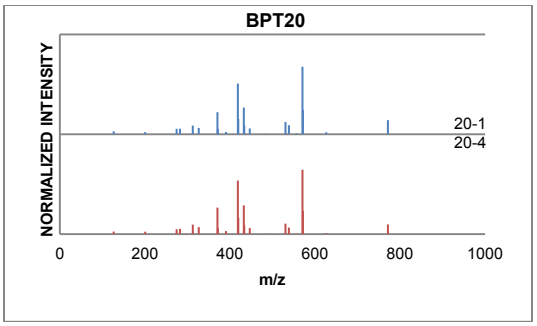
Averaged Mass Spectrum for BPT17



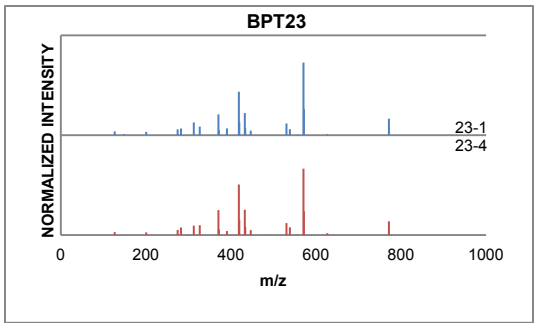
Averaged Mass Spectrum for BPT18



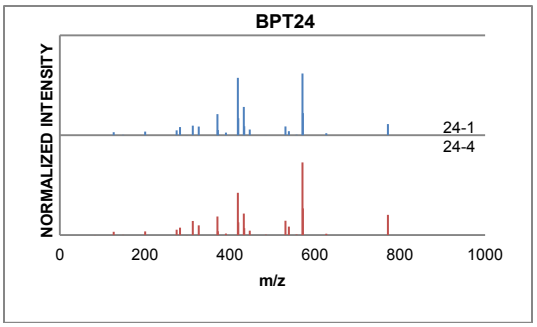
Averaged Mass Spectrum for BPT19



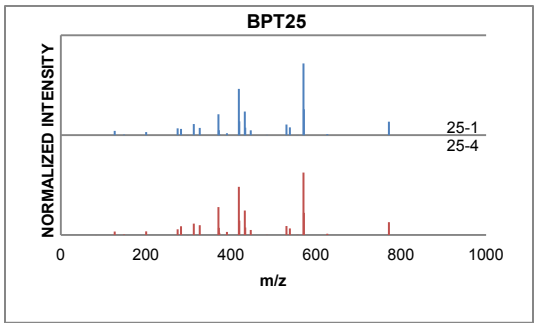
Averaged Mass Spectrum for BPT20



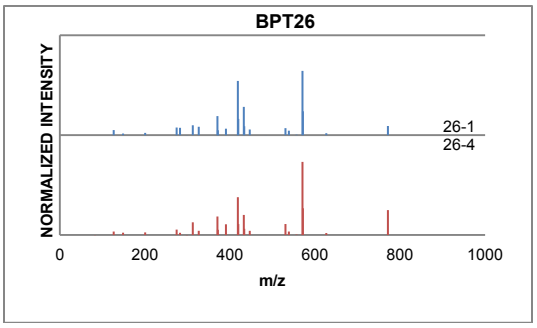
Averaged Mass Spectrum for BPT23



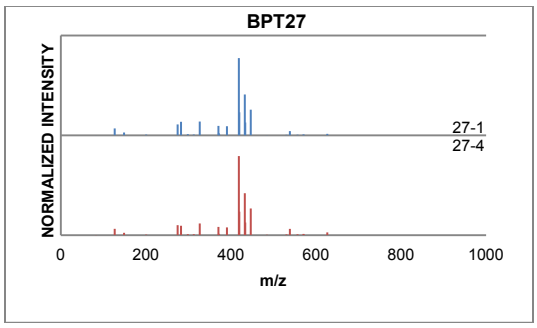
Averaged Mass Spectrum for BPT24



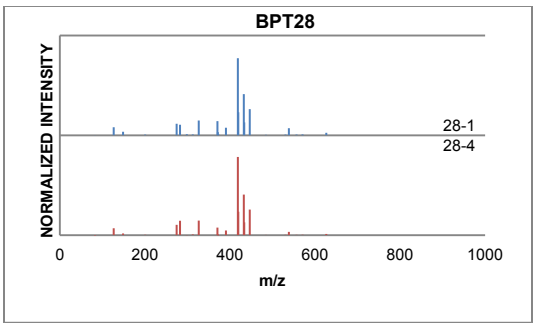
Averaged Mass Spectrum for BPT25



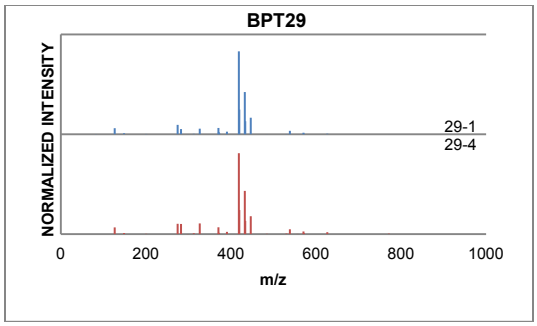
Averaged Mass Spectrum for BPT26



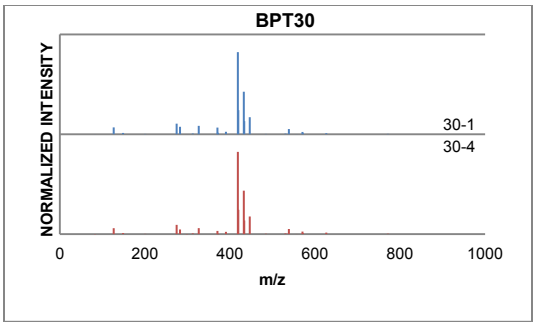
Averaged Mass Spectrum for BPT27



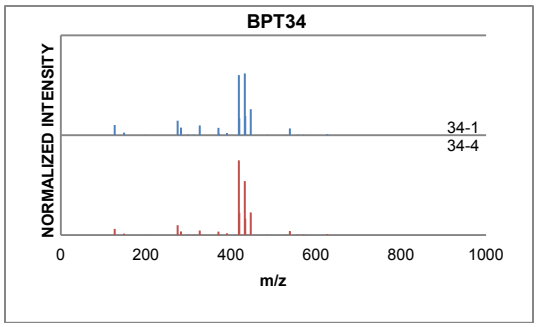
Averaged Mass Spectrum for BPT28



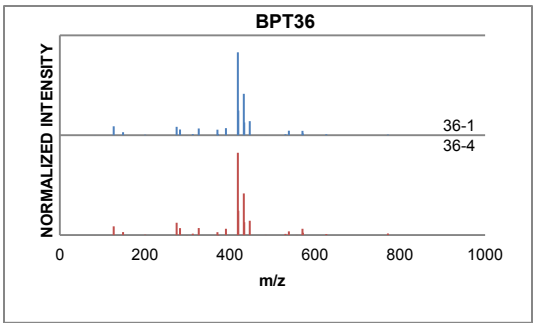
Averaged Mass Spectrum for BPT29



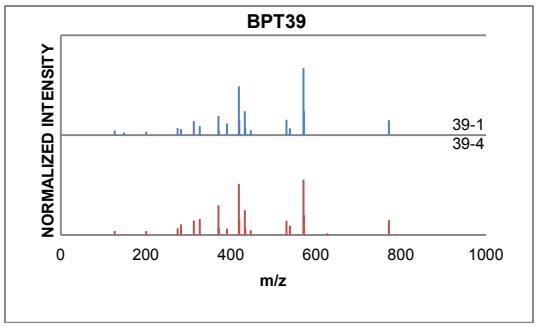
Averaged Mass Spectrum for BPT30



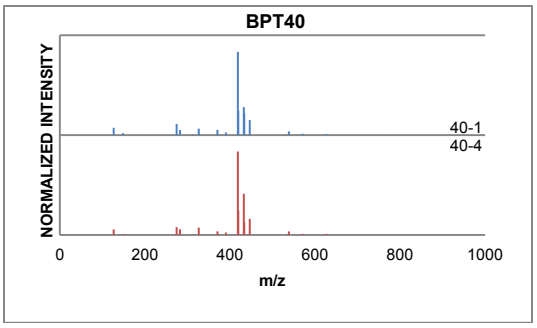
Averaged Mass Spectrum for BPT34



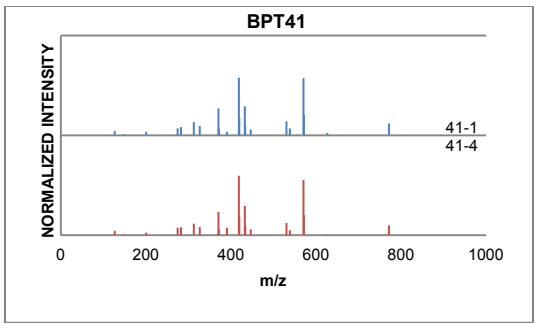
Averaged Mass Spectrum for BPT36



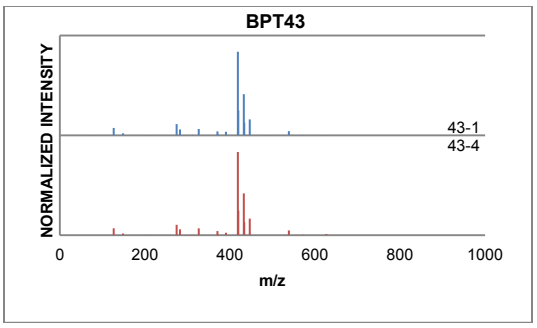
Averaged Mass Spectrum for BPT39



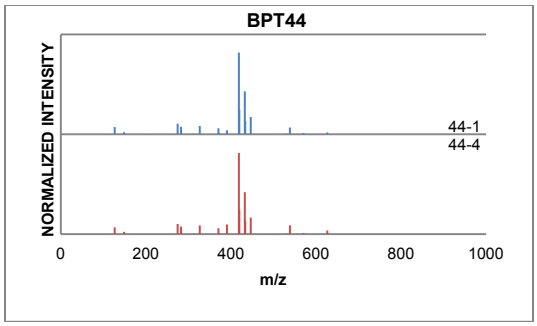
Averaged Mass Spectrum for BPT40



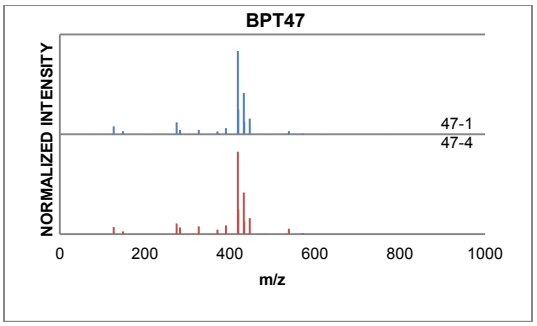
Averaged Mass Spectrum for BPT41



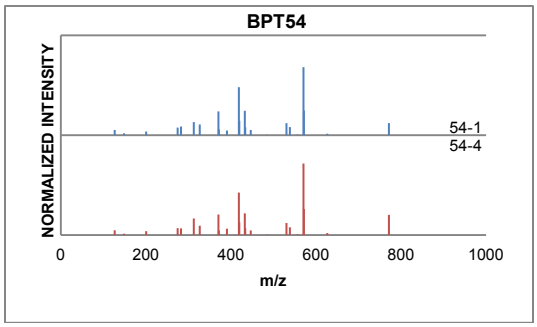
Averaged Mass Spectrum for BPT43



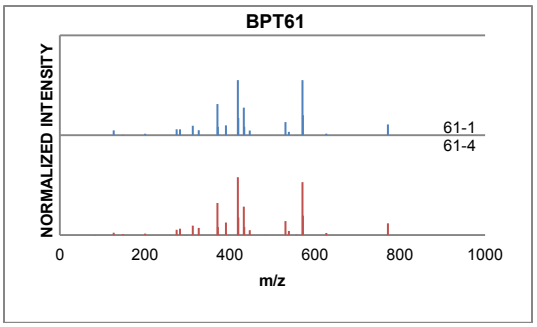
Averaged Mass Spectrum for BPT44



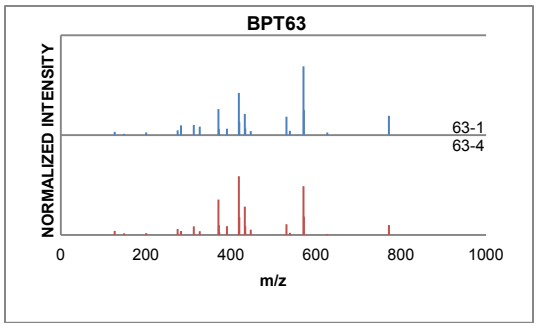
Averaged Mass Spectrum for BPT47



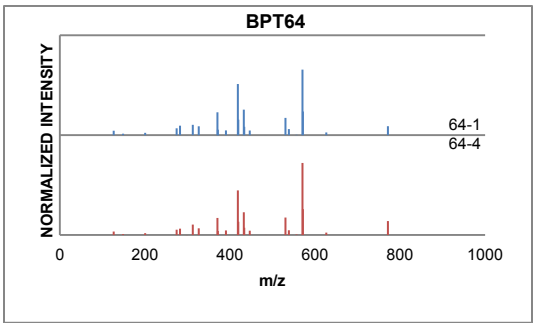
Averaged Mass Spectrum for BPT54



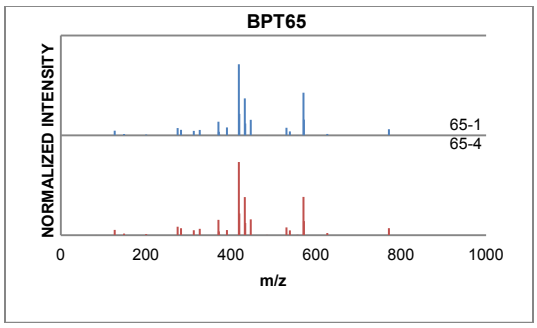
Averaged Mass Spectrum for BPT61



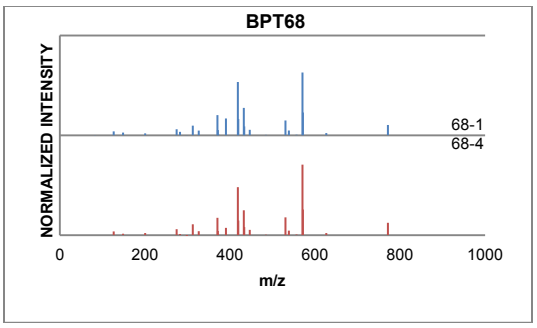
Averaged Mass Spectrum for BPT63



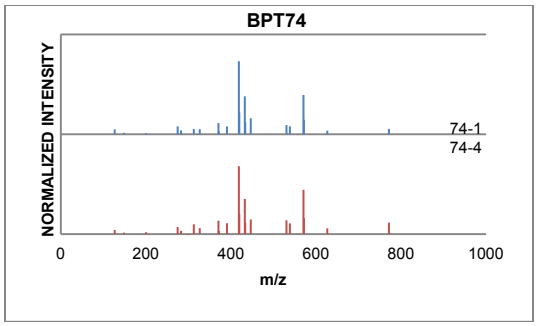
Averaged Mass Spectrum for BPT64



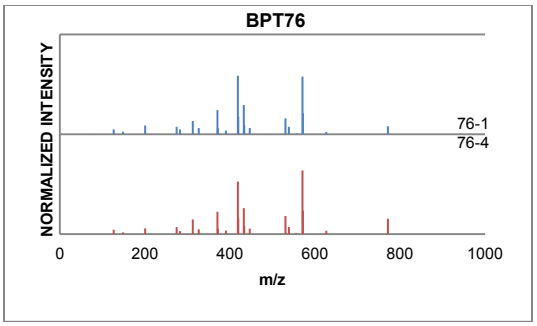
Averaged Mass Spectrum for BPT65



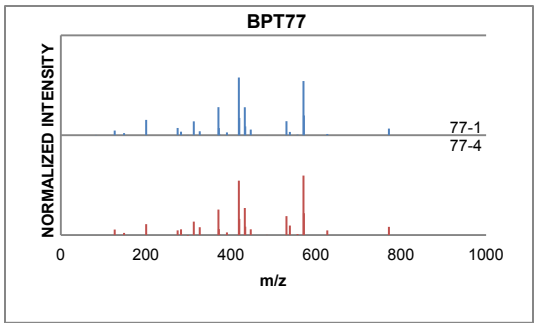
Averaged Mass Spectrum for BPT68



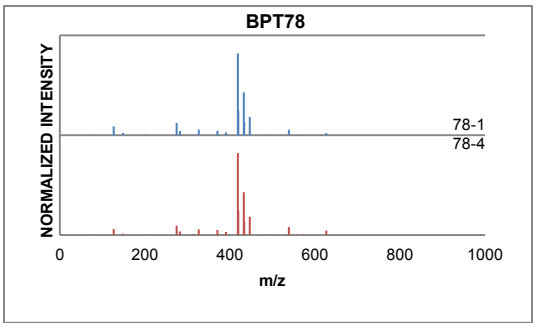
Averaged Mass Spectrum for BPT74



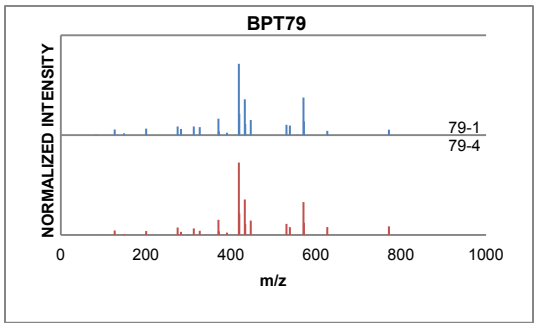
Averaged Mass Spectrum for BPT76



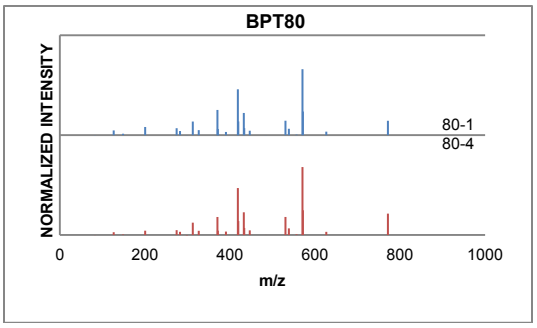
Averaged Mass Spectrum for BPT77



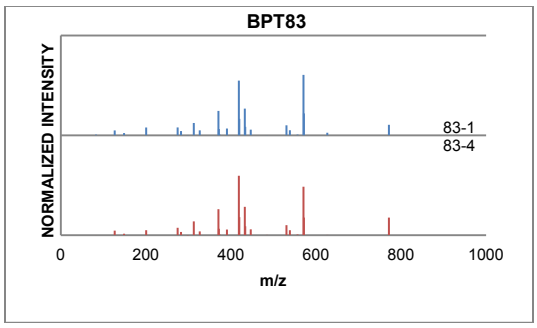
Averaged Mass Spectrum for BPT78



Averaged Mass Spectrum for BPT79

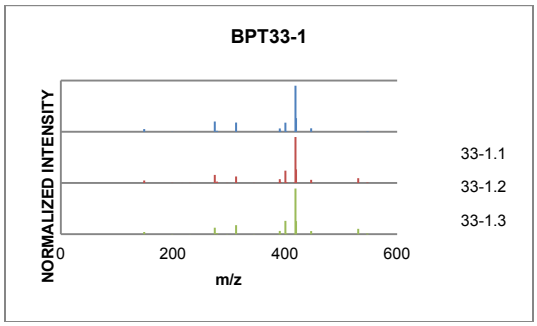


Averaged Mass Spectrum for BPT80

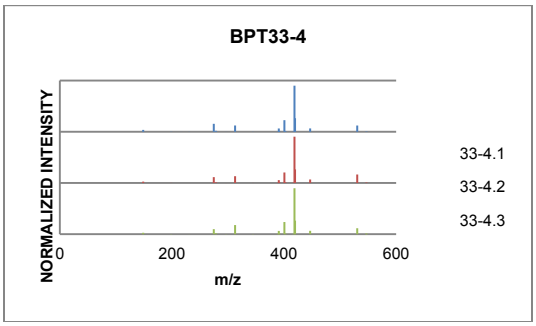


Averaged Mass Spectrum for BPT83

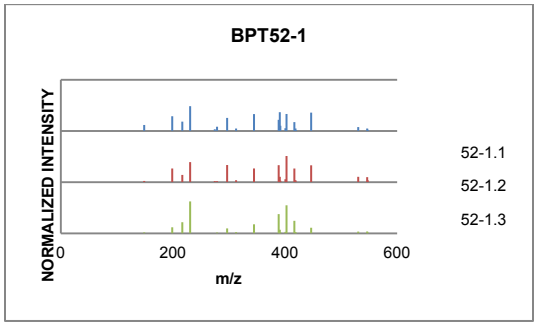
B.2.3. Tinted Adhesives



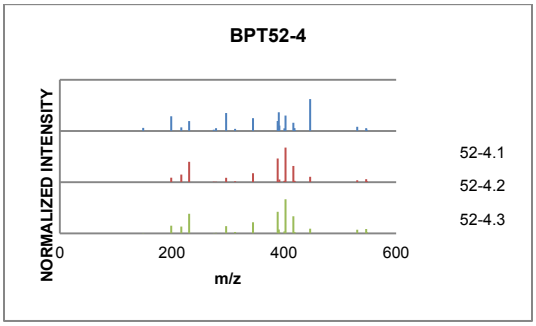
Mass Spectrum for BPT33-1



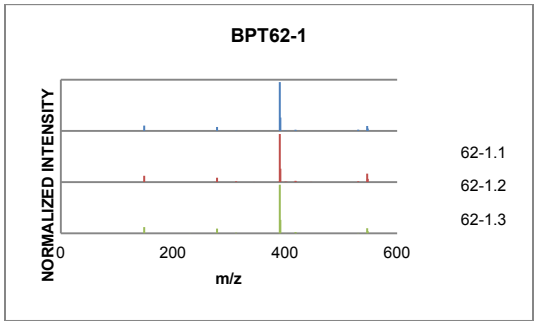
Mass Spectrum for BPT33-4



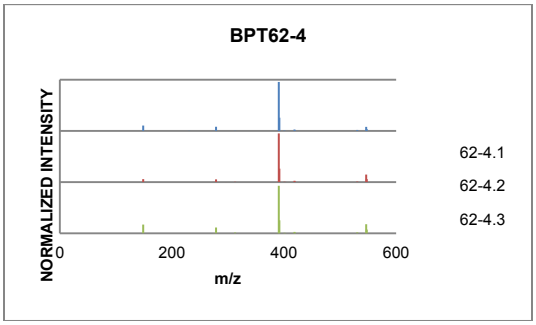
Mass Spectrum for BPT52-1



Mass Spectrum for BPT52-4

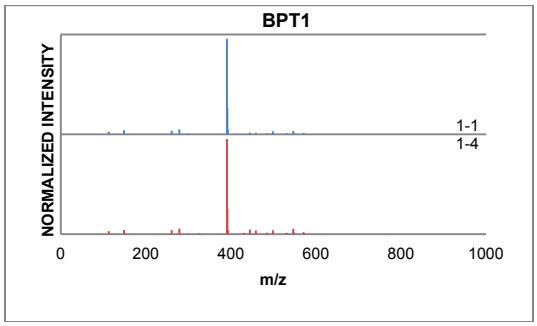


Mass Spectrum for BPT62-1

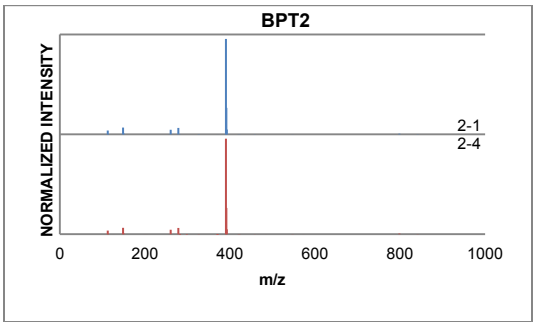


Mass Spectrum for BPT62-4

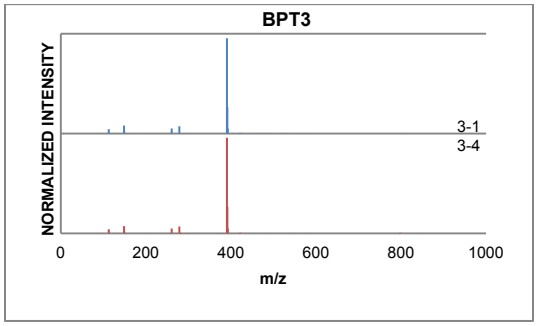
B.2.4. Colorless Adhesives



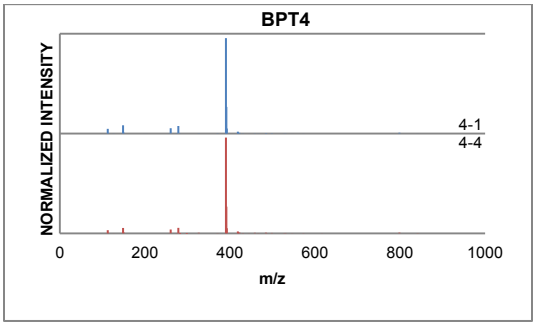
Averaged Mass Spectrum for BPT1



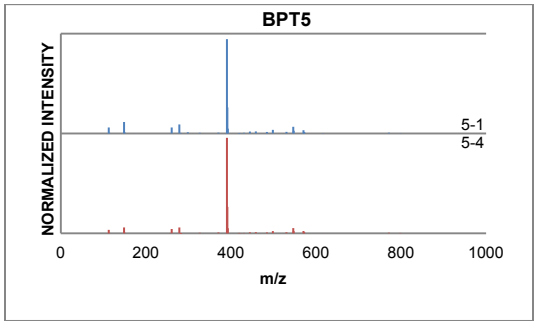
Averaged Mass Spectrum for BPT2



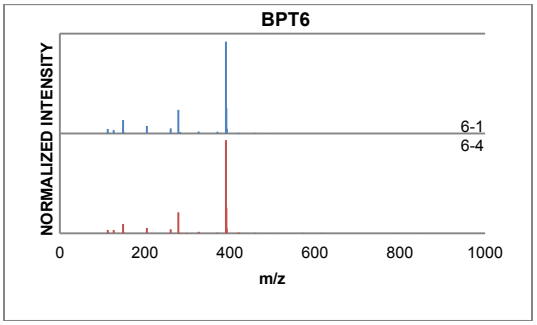
Averaged Mass Spectrum for BPT3



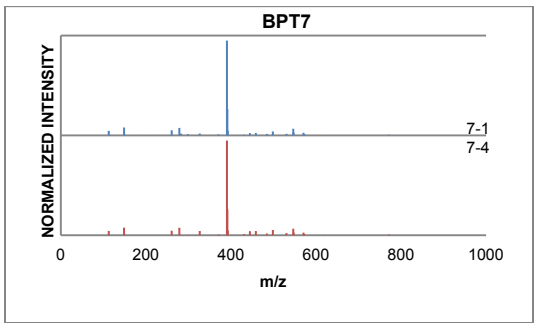
Averaged Mass Spectrum for BPT4



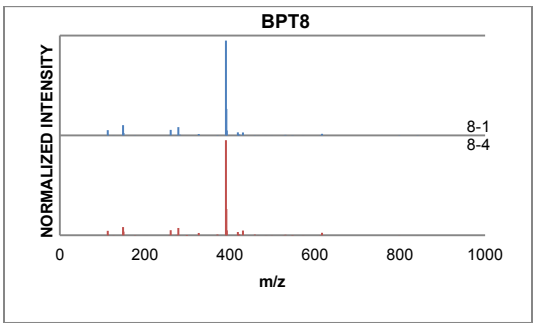
Averaged Mass Spectrum for BPT5



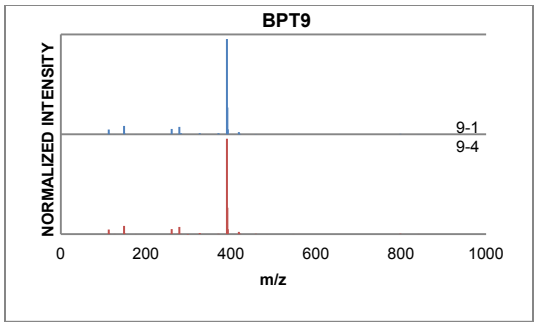
Averaged Mass Spectrum for BPT6



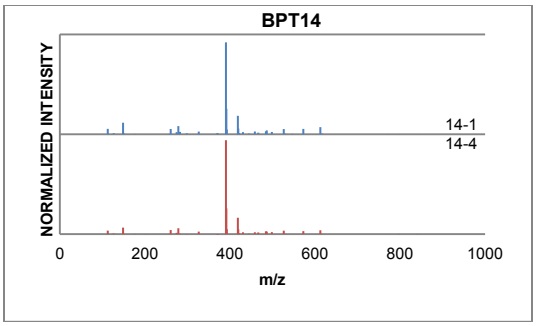
Averaged Mass Spectrum for BPT7



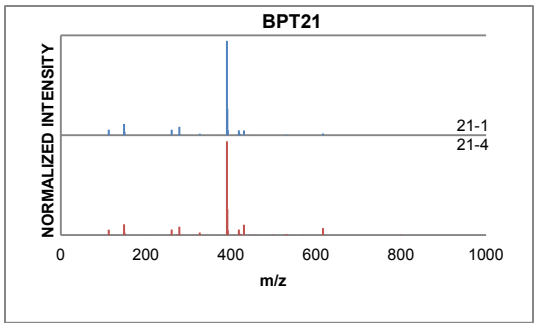
Averaged Mass Spectrum for BPT8



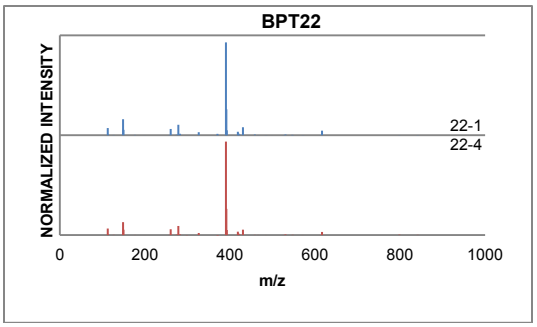
Averaged Mass Spectrum for BPT9



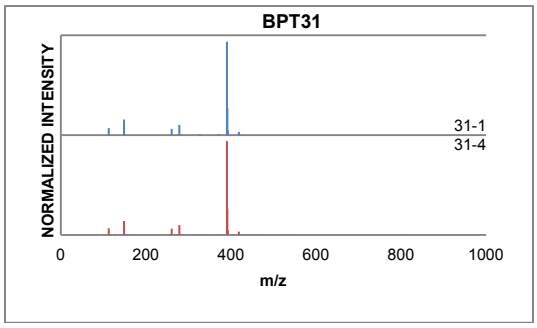
Averaged Mass Spectrum for BPT14



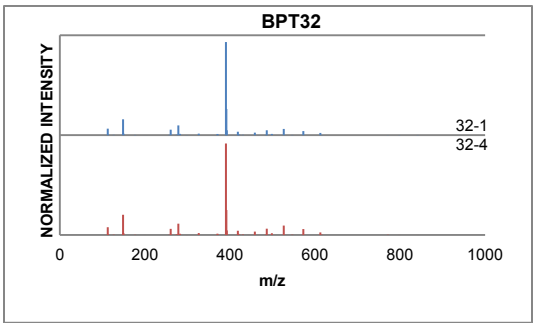
Averaged Mass Spectrum for BPT21



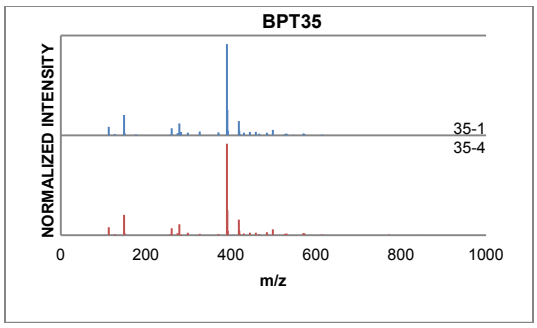
Averaged Mass Spectrum for BPT22



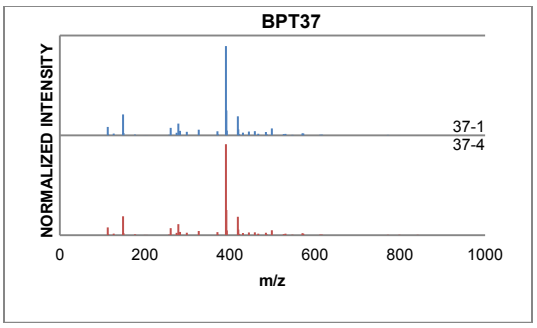
Averaged Mass Spectrum for BPT31



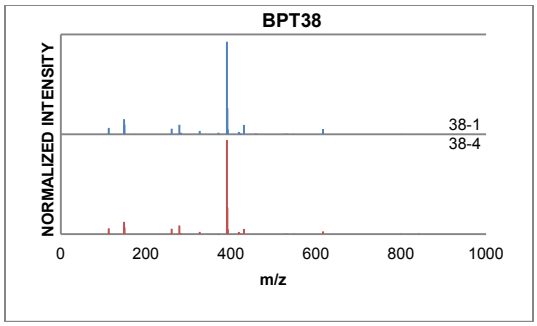
Averaged Mass Spectrum for BPT32



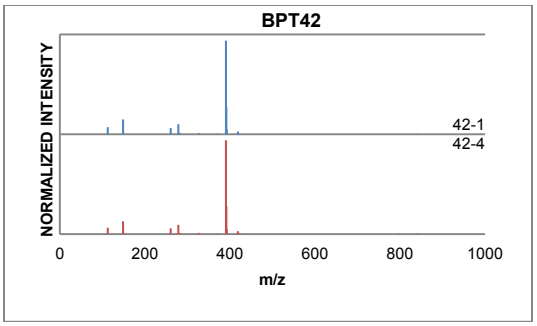
Averaged Mass Spectrum for BPT35



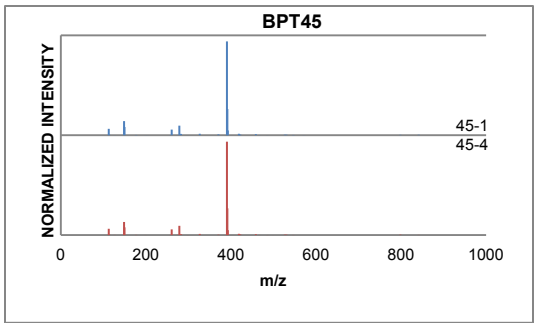
Averaged Mass Spectrum for BPT37



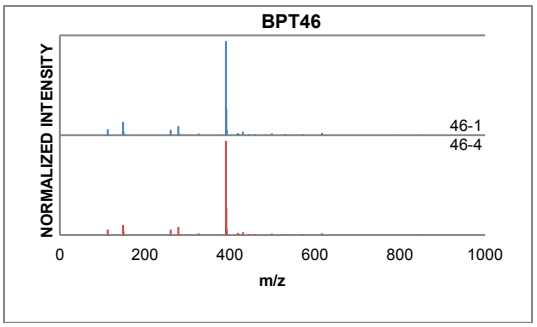
Averaged Mass Spectrum for BPT38



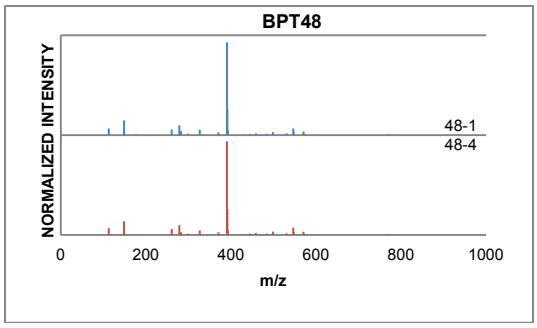
Averaged Mass Spectrum for BPT42



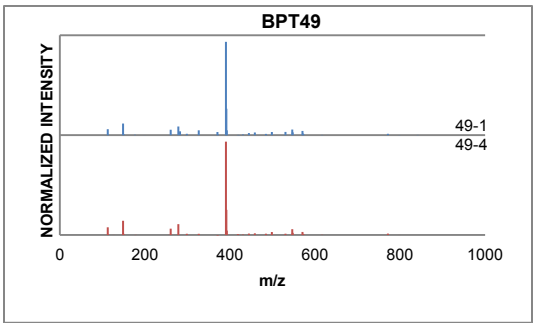
Averaged Mass Spectrum for BPT45



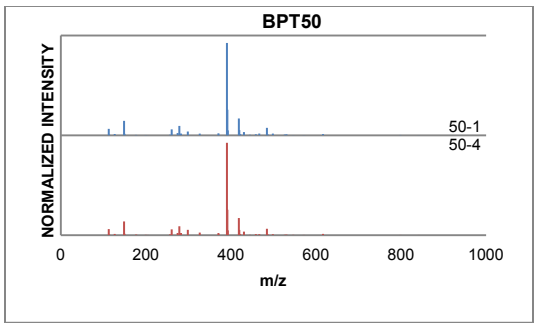
Averaged Mass Spectrum for BPT46



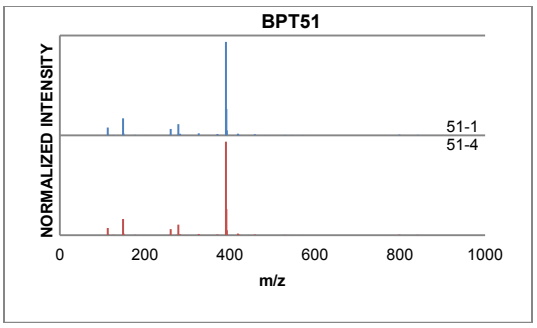
Averaged Mass Spectrum for BPT48



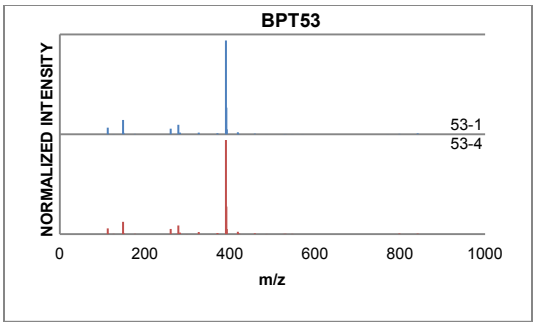
Averaged Mass Spectrum for BPT49



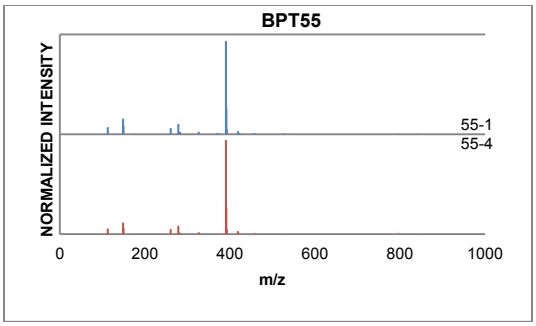
Averaged Mass Spectrum for BPT50



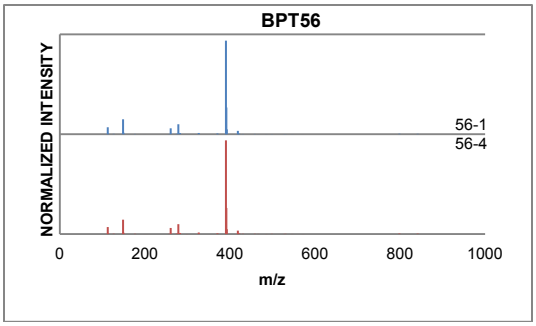
Averaged Mass Spectrum for BPT51



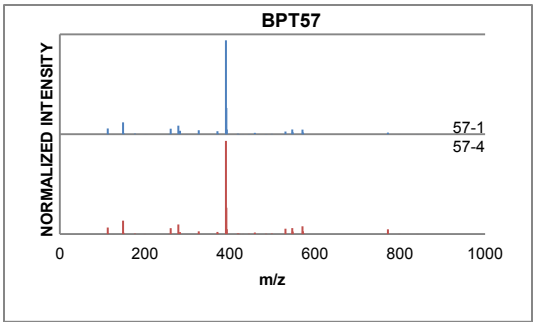
Averaged Mass Spectrum for BPT53



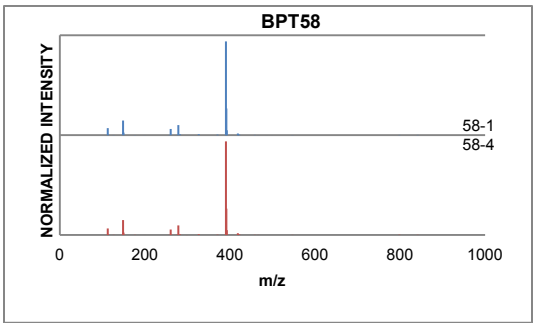
Averaged Mass Spectrum for BPT55



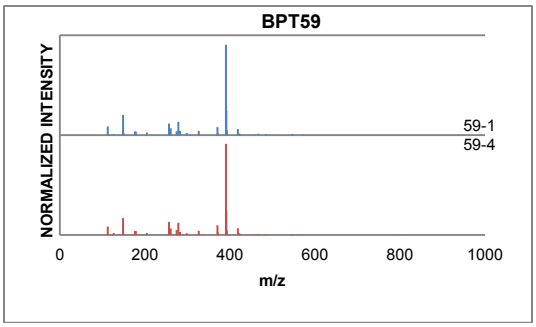
Averaged Mass Spectrum for BPT56



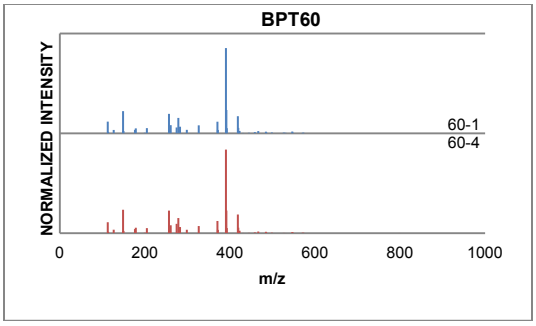
Averaged Mass Spectrum for BPT57



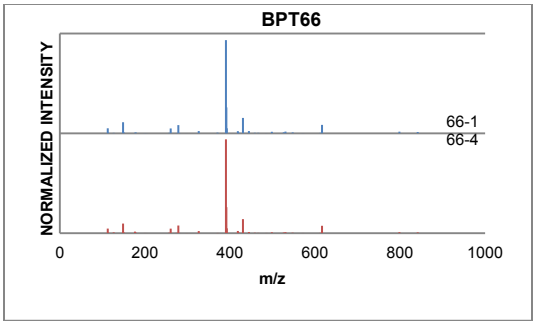
Averaged Mass Spectrum for BPT58



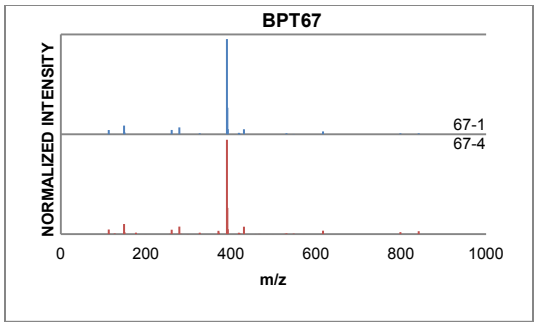
Averaged Mass Spectrum for BPT59



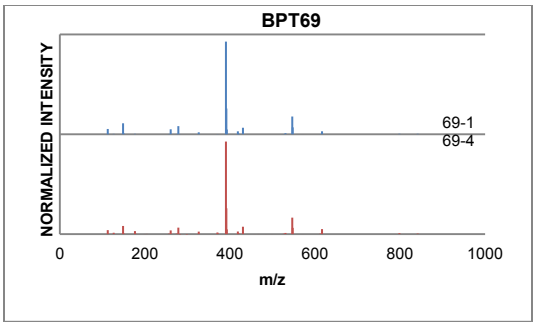
Averaged Mass Spectrum for BPT60



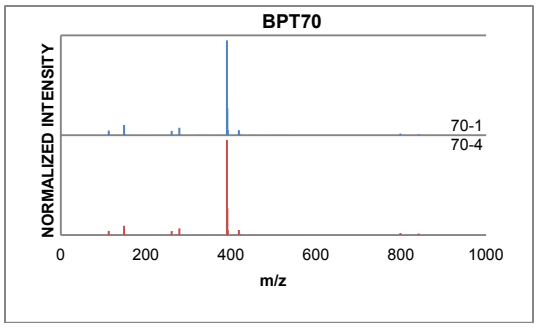
Averaged Mass Spectrum for BPT66



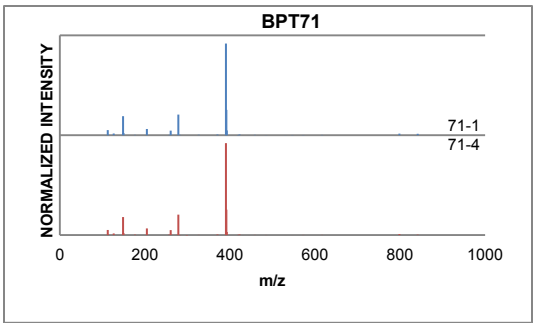
Averaged Mass Spectrum for BPT67



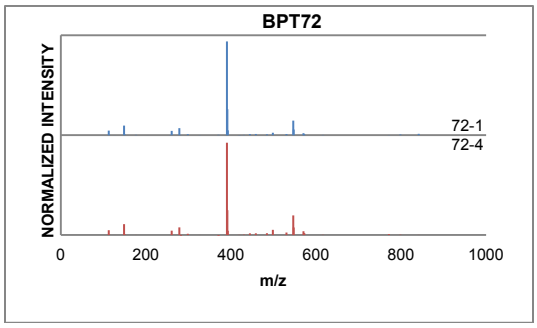
Averaged Mass Spectrum for BPT69



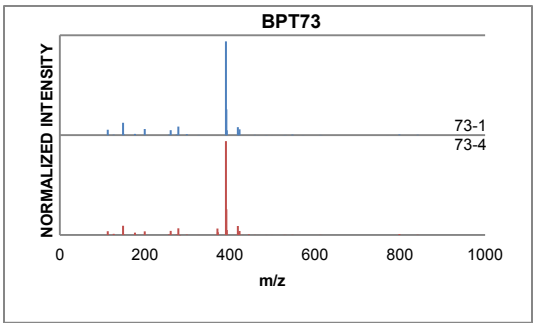
Averaged Mass Spectrum for BPT70



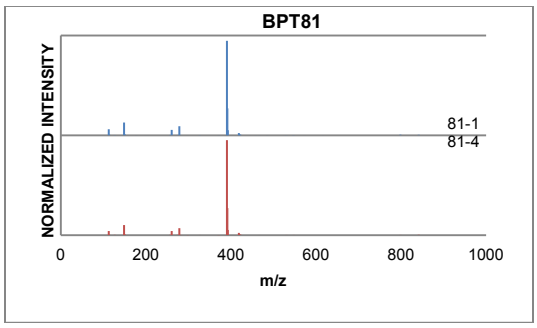
Averaged Mass Spectrum for BPT71



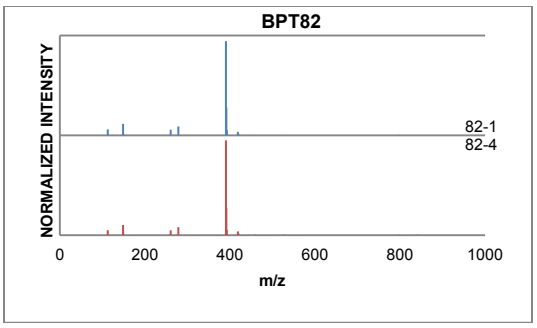
Averaged Mass Spectrum for BPT72



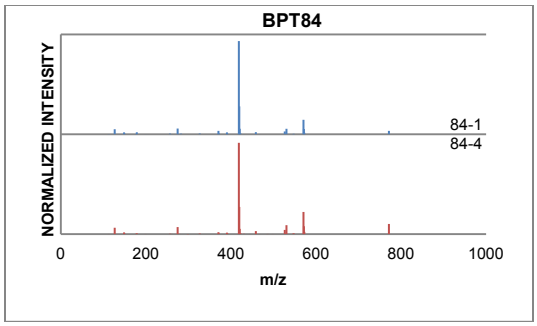
Averaged Mass Spectrum for BPT73



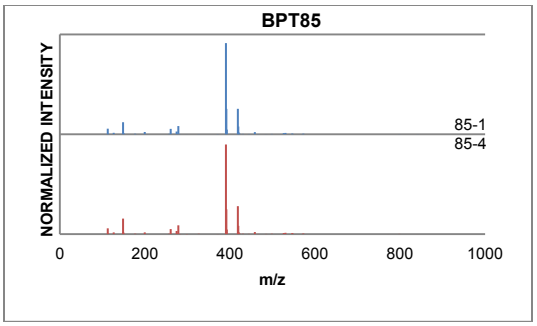
Averaged Mass Spectrum for BPT81



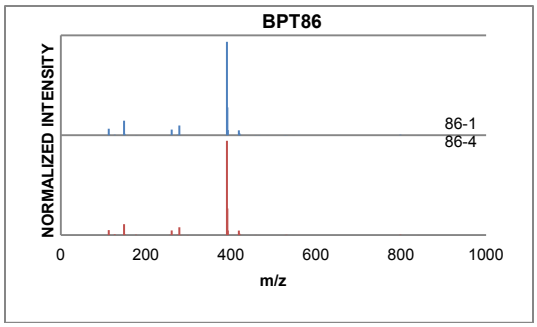
Averaged Mass Spectrum for BPT82



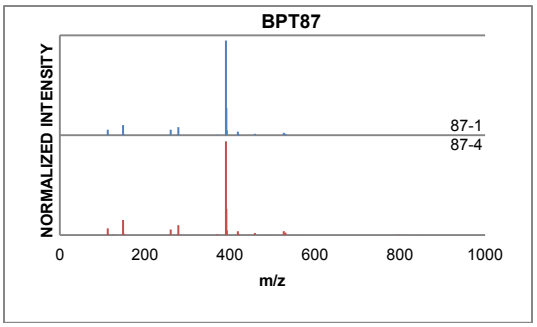
Averaged Mass Spectrum for BPT84



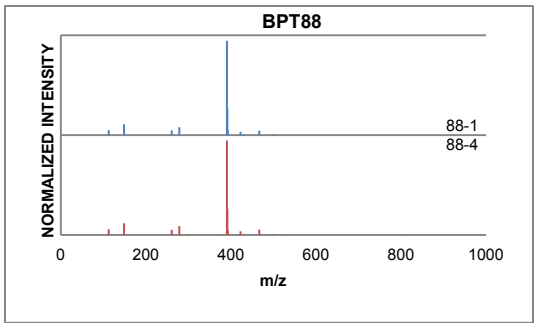
Averaged Mass Spectrum for BPT85



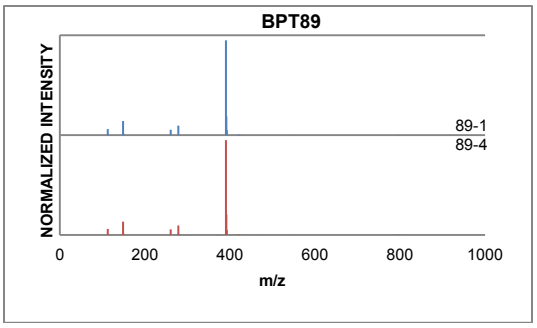
Averaged Mass Spectrum for BPT86



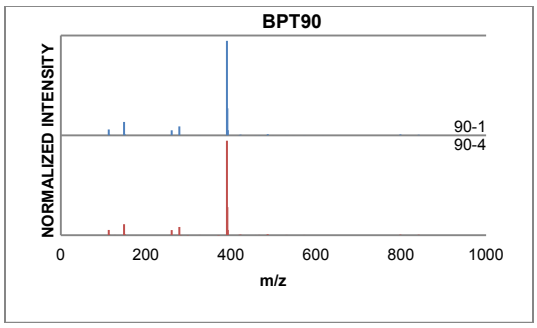
Averaged Mass Spectrum for BPT87



Averaged Mass Spectrum for BPT88



Averaged Mass Spectrum for BPT89



Averaged Mass Spectrum for BPT90